Link Transition Probabilities After a Local Strand Exchange in a Self-Avoiding Polygon Model of Enzyme-DNA Interactions

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Division of Biomedical Engineering of
University of Saskatchewan
Saskatoon

By
Marla A. Cheston

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Division of Biomedical Engineering
Room 2B60, College of Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5A9

Abstract

A new Local Strand Exchange (LSE) model is presented in this thesis as a first-step towards modelling recombinase-DNA interactions. The model uses self-avoiding polygons (SAPs) on the simple cubic lattice to represent DNA configurations. The new LSE model is related to the Local Strand Passage (LSP) model that was developed by Szafron and Soteros to model topoisomerase-DNA interactions.

The thesis begins with a review of the biological background for the topoisomerase and recombinase enzymes, followed by a review of current approaches to model the interactions of these enzymes with DNA. A review is then presented for the LSP model of a type II topoisomerase-DNA interaction, where a fixed structure, called Θ, is used as the interaction location. Following this, the new LSE model is introduced for modelling a site-specific recombinase-DNA interaction, where the Θ structure is again used as the interaction location. A recombination action is modelled with two different structures to take into account the two biologically relevant types of site-specific recombination. Specifically: 1) Direct repeat recombination sites are modelled with the direct-repeat-to-link strand exchange (DLE) structure; 2) Inverted repeat recombination sites are modelled with the inverted-repeat-to-knot strand exchange (IKE) structure.

The LSE model is studied using composite Markov chain Monte Carlo data that was previously generated to study the LSP model; the LSE study involves estimating link transition probabilities, *i.e.* the probabilities of going from knot type K to link type L after a strand exchange via DLE or IKE. Strong numerical evidence is provided that the link transition probabilities from the LSE model have asymptotic (as polygons get larger) properties consistent with conjectures based on polymer scaling theory. For example, it is shown that as polygon lengths increase, it becomes less likely that the fixed Θ structure will interact with any pre-existing knot. This means that after a strand exchange, typically, the original knot remains intact and a secondary knot or link can be formed at the Θ structure. In the DLE case, a link is created with one component of the link still containing the original knot, and in the IKE case, a composite knot containing the original knot can be created.

Lastly, for a subset of all the sampled polygons, numerical results that are biologically relevant are presented. Specifically, when the Θ structure is in the middle of a polygon (*i.e.* each of the two walks connected to Θ have the same length), then the most probable outcome is that the complexity of the knot will be reduced. In the DLE case, this is consistent with the mathematical and biological unlinking pathway observed in recombinase experiments. In the IKE case, the knot is more likely to become simpler than more complicated, reflecting recombinases's unknotting potential. This evidence establishes the new LSE model as a useful model for studying recombinase-DNA interactions, and worthy of future research.

Acknowledgements

Most importantly, I would like to thank my supervisors Dr. Chris Soteros and Dr. Michael Szafron. Without their knowledge, support, patience, and guidance this thesis would not have been possible. I would also like to thank my examining committee, Dr. Gordon Sarty, Dr. Ian Stavness, and Dr. Kaori Tanaka, for taking the time to read my thesis and for their helpful suggestions and comments. I extend my thanks to all the professors I've taken classes from throughout my University career. I would like to thank Michael Szafron again for allowing me access to his computer code which was immeasurably helpful, and Dr. R. Sharein for allowing me to use KnotPlot to run analysis and create many of my figures.

I would like to thank NSERC and the Division of Biomedical Engineering for their financial support throughout my program. Access to Compute Canada's WestGrid allowed me to run, and re-run, my programs much quicker and more efficiently than otherwise would have been possible.

I would like to thank my family for their unending support, as well as my friends who helped me retain my sanity throughout my university career. Lastly I would like to thank Jeremy Eng, Matthew Schmirler and Kevin McGregor for their helpful computer code, discussions, understanding of 'the process', and friendship.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

LSP	Local Strand Passage
LSE	Local Strand Exchange
DNA	Deoxyribonucleic Acid
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
bp	Base Pairs
E. coli	Escherichia coli
TopoIV	Topoisomerase IV
XerCD recombination	Xer recombination at XerC and XerD sites
FtsK_N	Domain one, the N-terminal domain of FtsK
FtsK_C	Domain three, the C-terminal domain of FtsK
FtsK_{50C}	FtsK derivative with the third domain intact
XerCD-*dif*	XerCD recombination at *dif* sites
XerCD-*dif*-FtsK	XerCD recombination at *dif* sites with FtsK present
SAW	Self-avoiding walk
SAP	Self-avoiding polygon
DLE	Direct-Repeat-to-Link Strand Exchange
IKE	Inverted-Repeat-to-Knot Strand Exchange
SP	Strand Passage
SE	Strand Exchange
MCMC	Markov Chain Monte Carlo
BFACF	Berg, Forester, Aragão de Carvalho, Caracciolo and Fröhlich
Θ-BFACF	BFACF algorithm account for Θ
CMCMC	Composite Markov Chain Monte Carlo
CMC Θ-BFACF	Composite Markov Chain Θ-BFACF algorithm
G.O.F.	Goodness of Fit

List of Symbols

K	A knot
L	A link
ϕ	An unknot
3_1	A trefoil knot
4_1	A figure eight knot
3_1^+	A positive trefoil knot
3_1^-	A negative trefoil knot
$K_1 \# K_2$	Knot K_1 composed with knot K_2
0_1^2	A unlink link
2_1^2	A Hopf link
$\|\omega\|$	Number of edges in SAW or SAP ω
c_n	Number of distinct n-edge SAWs
p_n	Number of distinct n-edge SAPs
$p_n(K)$	Number of n-edge SAPs with knot type K
$k(\omega)$	Knot type of ω
$\frac{p_n(K)}{p_n}$	Probability that SAP has particular knot type K
κ	Connective constant for SAW or SAP
μ	Growth constant for SAW or SAP
Θ	The fixed structure for LSP model
$\Theta - SAP$	A SAP which contains the Θ structure
Θ_0	Θ structure which would have successful strand passage
Θ_S	The after-strand-passage Θ structure
p_n^{Θ}	Number of distinct n-edge Θ-SAPs
$p_n^{\Theta_0}$	Number of distinct n-edge Θ_0-SAPs
$p_n^{\Theta_S}$	Number of distinct n-edge Θ_S-SAPs
$p_n^{\Theta}(K)$	Number of distinct n-edge Θ-SAPs with knot type K.
$p_n^{\Theta_0}(K)$	Number of distinct n-edge Θ_0-SAPs with knot type K.
$\rho_n^{\Theta_0}(K \to K')$	Knot transition probability from knot type K to K'
$\rho_n^{\Theta_0}(K \to s)$	Successful strand passage probability from knot type K
Θ_1	Θ structure which would have successful strand exchange
Θ_{DLE}	The after direct-to-link strand exchange Θ structure
Θ_{IKE}	The after inverted-to-knot strand exchange Θ structure
$\Theta_1 - SAP$	A SAP which contains the Θ_1 structure
$p_n^{\Theta_1}$	Number of distinct n-edge Θ_1-SAPs
$p_n^{\Theta_1}(K)$	Number of distinct n-edge Θ_1-SAPs with knot type K
$p_n^{\Theta_{II}(K)}$	Number of n-edge class II Θ-SAPs with knot type K
$\rho_n^{SE}(K \to L)$	Link transition probability from knot type K to link type L
$\rho_n^{SE}(K \to s)$	Successful strand exchange probability from knot type K
$(+1)$	Tangle analysis which adds a positive crossing
(-1)	Tangle analysis which adds a negative crossing
(0)	Tangle analysis where no crossing is introduced
$\{X_t, t \in T\}$	A stochastic process
$\mathcal{S}$	The state spaces of a stochastic process
P_{xy}	Probability of going from state x to state y in one step
$P_{xy}^{(n)}$	Probability of going from state x to state y in n steps
$p_n(+2)$	A plus two BFACF move
$p_n(-2)$	A minus two BFACF move
$p_n(0)$	A zero BFACF move
z	The fugacity of a chain

$z_c(K)$	The critical z value for knot type K
$SE(g)$	Standard error of variable g
$\pi_\omega(q, z)$	Equilibrium distribution for BFACF or Θ-BFACF algorithm
$\widehat{\delta}_n^r(K)$	Estimated relative standard error of variable
$\widehat{\delta}^r(K)$	Minimum estimated relative standard error over all n
ϵ	Acceptable tolerance of $\widehat{\delta}_n^r(K)$ for variable
$N_{max}(K)$	The polygon length, n, where data become unreliable for knot type K
τ_{exp}	The exponential autocorrelation time
$\tau_{exp,W}$	τ_{exp} estimate found using warm-up analysis
$\tau_{exp,E}$	τ_{exp} estimate found using potential scale reduction method
$\sqrt{\hat{R}_j}$	Elements of the sequence are the estimated potential scale reduction
$\tau_{int,A}$	The integrated autocorrelation time for function A
$t_{K,L}^{SE}$	Asymptotic value estimate for *Case 1* from knot type K to link type L
$C_{K,L}^{SE}$	Asymptotic value estimate for *Case 2* from knot type K to link type L
$\rho_n^{SE*}(K \to L)$	$\rho_n^{SE}(K \to L)$ for *Case 2* when *Case 1*'s influence has been eliminated
Λ	Goodness of fit test statistic
$\Gamma_{SE,C1}$	Knot not containing the original knot for *Case 1*
equal-length	Θ-SAPs where Θ is connected with two equal-length SAWs
'equal-length'	Θ-SAPs where Θ is connected with two SAWs with lengths within 24-edges of each other
variable-length	Θ-SAPs where Θ is connected with two SAWs of any length

Chapter 1

Introduction

The focus of this thesis is on modelling the knotting and linking simplification role of enzyme-DNA interactions, particularly modelling the interaction of two types of enzymes, topoisomerase and recombinase, with DNA. These enzymes are of interest because they have been shown to play essential biological roles in all cells [12, 16]. For example, for a cell to survive, its DNA must be replicated, but in order for replication to take place in circular DNA, the circular DNA must be unlinked and unknotted. Type II topoisomerase is the main enzyme responsible for unlinking and unknotting circular DNA; however, it has recently been shown that, when topoisomerase is inhibited, site-specific recombinase is able to take its place, allowing the cell to successfully replicate [5, 47, 53, 95]. In order to model these enzyme-DNA interactions, the existing Local Strand Passage (LSP) model of type II topoisomerase action [105, 106], will be reviewed, and a newly created Local Strand Exchange (LSE) model will be introduced to model site-specific recombinase action. Each action will be performed on self-avoiding polygons (SAPs) on the simple cubic lattice.

Chapter 1 will provide the introduction for the thesis by first reviewing the biological background on topoisomerases and site-specific recombinases, and motivating their study. The focus will be on site-specific recombination performed by recombinase enzymes XerC and XerD at *dif* sites (XerCD-*dif*); these are the recombinases responsible for unlinking and unknotting in the commonly studied *Escherichia coli* (*E. coli*). Although the focus is on specific enzymes, many of the topics covered in this thesis apply to a more general enzyme-DNA interaction model. Secondly, in Section 1.2 an overview of knots will be covered, followed by a brief review of previous enzyme-DNA models. The subsequent chapter will then focus on reviewing the existing LSP model for the topoisomerase-DNA interaction and introducing the new LSE model for the recombinase-DNA interaction. The final chapters will present numerical studies of these models using Markov chain Monte Carlo methods. Specifically, using these models, knotting and linking simplification will be investigated through estimates of link transition probabilities. Through link transition probability estimates, it will be shown that site-specific recombinases, modelled with the LSE model, have the ability to change the knot and link type of DNA. Additionally, for a biologically relevant subset of polygons, evidence will be presented indicating that a simplification of topology (knot or link type) is most probable, and this evidence is consistent with DNA experiments.

1.1 Biological Background

This section will first provide an overview of DNA, and the descriptions of the action of type II topoisomerases and site-specific recombinases on DNA. Type II topoisomerases will be shown to be the main enzymes responsible for unknotting and unlinking DNA, allowing cells to properly replicate and survive (for precise definitions of knots and links see Section 1.2). Site-specific recombinases will then be introduced, focusing on XerCD recombinase at *dif* sites. It will be shown that in the presence of an additional enzyme, FtsK, Xer recombinases act alongside type II topoisomerases in order to unknot and unlink the DNA, or in the absence of type II topoisomerase in *E. coli*, Xer recombinases can unlink replication links. The section will close with a brief summary of the highlights from the biological background which are especially pertinent for the remainder of this work.

1.1.1 DNA Overview

Deoxyribonucleic acid, better known as DNA, is responsible for encoding genetic information in all living organisms on Earth [16]. DNA is composed of repeated units called nucleotides, which consist of three components: a phosphate group, a sugar, and one of four bases: adenine (A), guanine (G), thymine (T) or cytosine (C) [12]. DNA's now well-known double-stranded helical nature, was proposed by Watson and Crick, with the help of Franklin's work, in 1953 [40, 117]. The phosphates and sugars form the two backbones of DNA, and the bases pair either A with T, or C with G, joining the backbones together, *cf.* Figure 1.1. The two strands of DNA are called *complementary strands* since if one strand has A the other will have the complementary T in the corresponding location. The two strands twist around each other in a right-handed, helical fashion, *cf.* Figure 1.1. The length of DNA widely varies: the human genome is made of approximately 3 billion base pairs (bp); *Escherichia coli* (*E. coli*), one of the most studied bacteria, has 4.4 million bp [16]; and plasmids, which are small genetic elements separate from chromosomal DNA in micro-organisms [36], typically have about 5000 bp [84].

DNA can be found in a linear form, such as in human and other mammalian DNA, or it can be circular, such as in bacterial DNA, chloroplast DNA, and human mitochondrial DNA [16]. In circular DNA a closed double-stranded circle is formed, *cf.* Figure 1.1. Although the two types of DNA have different properties, linear DNA is often anchored to a protein scaffolding a number of times within the cell [12]. This anchoring to the scaffolding restricts the rotation of the DNA and fixes the ends of the DNA relative to each other, so in a topological sense, linear DNA can behave as closed circular DNA [12, 116].

On a larger scale, DNA can be viewed as an example of a copolymer. A linear homopolymer is a chain-like molecule consisting of many repeated molecular units, called monomers, joined together with chemical bonds, while a linear copolymer is a molecular chain formed from more than one type of monomer [65]. A ring polymer, or copolymer, is one where the ends are attached together. In the case of DNA, the base pairs can be considered as two different monomer types (A-T and C-G), so they form a copolymer when they are

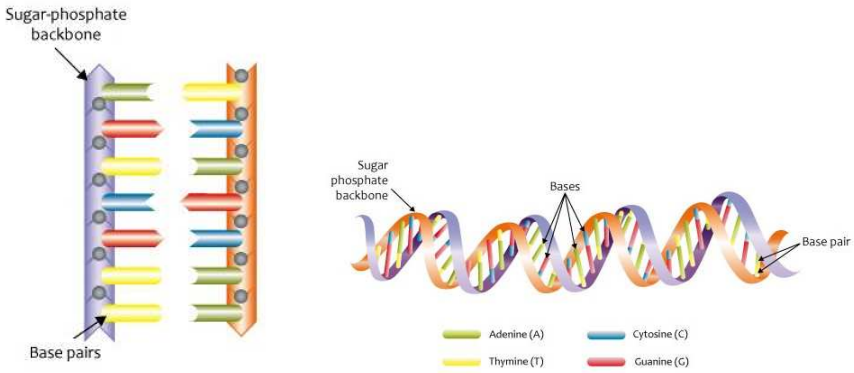


Figure 1.1: Structure of DNA. The left figure shows the sugar-phosphate backbone of the DNA with the four base pairs joining the backbones together. The right figure shows the helical nature of DNA. Reproduced from [31, 32] in accordance with Creative Commons BY 2.0, edited by Cheston.

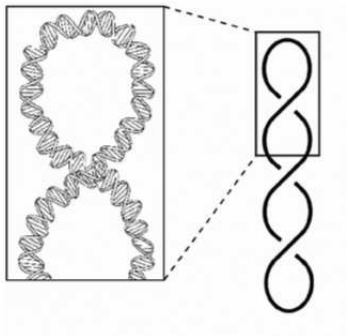


Figure 1.2: Supercoiling of DNA. The close up shows a representation of the double-stranded DNA helix as a single line focusing on the axis the DNA winds around. This main picture shows negative (by the right-hand rule) supercoiling of the DNA. Reproduced with permission of Springer New York in format Thesis/Dissertation via Copyright Clearance Center [39].

joined together. Therefore, on a macroscopic scale, circular DNA can be thought of as a ring copolymer when focusing on the axis the DNA winds around, *cf.* Figure 1.2. This will be how DNA will be viewed in this work.

An inevitable result of the double helical structure of DNA is supercoiling [12]. Supercoiling is the coiling of helical DNA around itself *cf.* Figure 1.2. Supercoiling plays an important role in cell processes, such as replication and transcription [28], and it allows DNA, which can be very long, to fit compactly in cells [12]. For example, *E. coli* DNA is 1.5 mm in length but must fit into a cell with a dimer of less than 1 μm, so supercoiling assists in the packing process [16]. *In vivo*, DNA supercoiling is negative, making for easier access to the base pairs, greatly assisting the replication process [16].

DNA can be knotted. Knots were first observed in single-stranded circular DNA in 1976 by Liu *et al.* [60], then in double-stranded DNA in 1981 by Liu *et al.* [61]. Since then, knotting has been researched *in vivo*, *in vitro* and *in silico*. Yan *et al.* showed that the most energetically favourable conformation *in vitro* for 200 kilo bp DNA molecules at thermal equilibrium is the trefoil knot [121]. Arsuaga *et al.* experimentally showed

that small volumes stimulate knotting, and *in silico* that the longer the genome, the higher the probability of knotting [3]. Supercoiling has also been shown to promote strand tangling by bringing distant strands into close proximity [28]. Despite these reasons for DNA to be knotted, circular DNA is not commonly expected to be knotted *in vivo* [12]. This is due to a number of reasons, including: 1) knotted DNA is specifically problematic for cells, because necessary cell processes, such as replication, cannot occur when DNA is knotted [12, 16]; 2) specific enzymes exist in order to remove knots (see Sections 1.1.2 and 1.1.3) and these have been found to be very efficient at unknotting and unlinking [86]; and 3) it has been shown that negatively supercoiled DNA has lower free energy than knotted DNA, so supercoiling suppresses knotting [63]. Knotted polymers also have reduced tensile strength; therefore knotted DNA is more likely to break than unknotted DNA [63].

The main type of enzymes responsible for resolving DNA knots and links, are type II topoisomerases. However, recently, a second type of enzymes, site-specific recombinases, have been found to also have the ability to unknot and unlink DNA [5, 47, 53, 95]. Topoisomerase, specifically type II topoisomerase, and recombinase, specifically Xer recombinase, will be discussed next.

1.1.2 Topoisomerase Action on DNA

In this subsection the two types of topoisomerase will be introduced, with a focus on type II topoisomerase. Type II topoisomerase's mechanism for unknotting and unlinking DNA, and the result of this mechanism, will be highlighted.

When DNA is replicated the original DNA, or chromosome, strands act as a template in order to create two new daughter strands of DNA [12]. During cell division, once the double-stranded daughter molecules have been created they soon begin to move away from the middle of the cell, or mid-cell, moving towards the poles of the cell in order to be *segregated* into two separate daughter cells (a general picture of this can be seen later in this section in Figure 1.14) [91, 10]. Once separated, the chromosomes are referred to as monomers (note monomers here refers to a single chromosome, not the previous molecular unit definition of monomers as part of a polymer). In order for the original DNA strands to act as a template for replication, they must first be separated; the 'Y'-position where the strands break apart is called the *replication fork*, *cf.* Figure 1.3. The separating of the two strands requires the helix to rotate ahead of the replication fork, causing positive supercoiling to build up ahead of the fork [47]; a representation of this can be seen in Figure 1.3. In order for successful replication to proceed, these supercoils must be removed. In 1971 James Wang discovered an enzyme called topoisomerase which reduced the supercoiling in DNA [12]. It turns out that topoisomerases exist in every type of cell and regulate supercoiling, linking and knotting in DNA [16]. Two separate families of topoisomerases exist: type I, which is capable of breaking one strand of DNA, and type II, which is capable of breaking both strands of the DNA.

Type I topoisomerase's main job is regulating supercoiling by binding to the DNA, breaking one strand, passing the other strand of DNA through, and then resealing the break and releasing the DNA. This action

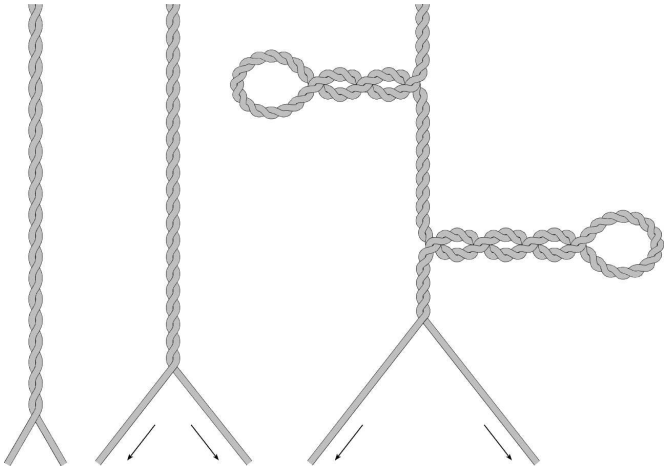


Figure 1.3: When replication takes place the strands of DNA must be pulled apart, creating a 'Y'-shaped replication fork. This untwisting of the strands adds additional positive supercoiling ahead of the replication fork. Reproduced by permission from Oxford University Press [12].

allows the supercoiling to be *relaxed*, preventing torsional strain on the DNA and allowing replication to continue [16]. Type II topoisomerase is capable of breaking both strands of DNA, and so is able to not only relax supercoiling, but also change the knot or link type of DNA. In *E. coli*, a type II topoisomerase called DNA gyrase, is responsible for removing the majority of the positive supercoils ahead of replication forks [12, 91, 94]. The main type of topoisomerase discussed here will be type II topoisomerases, particularly the type II topoisomerase in *E. coli* called topoisomerase IV.

Topoisomerase IV (topoIV), a type II topoisomerase in *E. coli*, relaxes DNA, but more importantly changes the knot or link type of the DNA [12]. TopoIV was discovered by Kato *et al.* in 1990, when they found that topoIV could relax DNA supercoils and that it was required in chromosome segregation [57]. TopoIV's mechanism can be seen in Figure 1.4. First topoIV binds to one double-stranded segment of DNA, called the gate segment, and then brings another segment, called the transported segment, through the gate. Once the transported segment is through, the gate segment is resealed and the topoisomerase releases the gate segment to move on and act elsewhere [70]. Because this operation passes one strand of DNA through another strand, it is called a *strand passage* (SP).

In vitro DNA experiments by Rybenkov *et al.* in 1997 showed that type II topoisomerases are very effective at unknotting and unlinking DNA [86], which is very important for the replication and segregation of the chromosomes. During replication, *E. coli* is replicated bidirectionally from a replication origin towards the termination region on the opposite side of the chromosome, meaning that there exist two replication forks, one in each direction (shown in summary Figure 1.18 in Section 1.1.4 with the black dots). When the forks converge in the termination region, topoisomerases, such as gyrase (a type II topoisomerase), can no longer resolve the supercoiling due to space restrictions. This supercoiling then diffuses into the newly replicated DNA behind the fork, where the replicated chromosomes become intertwined together, forming *precatenanes* [53, 63, 91, 94]. If not removed by the end of replication, the precatenanes are converted into catenated

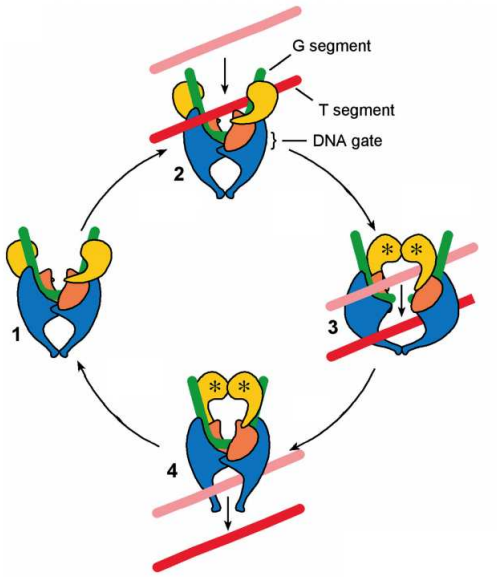


Figure 1.4: The mechanism of topoIV. TopoIV binds to the green gate (G) segment, then brings a second transported (T) segment through the gate. This action has the potential to change the knot type of the DNA and is called a strand passage. Reproduced from [11] under terms of Creative Commons Attribution-NonCommercial 2.5 Generic License, edited by Cheston.

monomers, or *cantenanes*, meaning the daughter chromosomes are now interlinked together [91]. *E. coli* has more than 400,000 links which must be removed in order for the daughter chromosomes to segregate properly and successfully replicate [94]. TopoIV is able to act and efficiently unlink the daughter chromosomes. If topoIV does not act, the daughter chromosomes will remain linked together as a catenane, and will not be able to divide properly, resulting in the cell's death [16]. TopoIV is also very efficient at unknotting DNA, in both a single knotted chromosome, or when multiple chromosomes are knotted together, *i.e.* when two chromosomes are bonded together into a single circle, becoming a *dimer*, which is then knotted [86]. TopoIV action on catenated chromosome monomers and a knotted dimer is illustrated in summary Figure 1.18 in Section 1.1.4 at A., B. and D.

Note that topoIV is not known to be site-specific (*i.e.* it does not bind to specific DNA base pair sequences) and determining what rules govern where it binds to the DNA in order to achieve such efficient unknotting and unlinking remains an open problem.

Due to their crucial role in many cell processes, including replication, topoisomerases are essential for all cells to survive [42, 70]. For this reason, topoisomerases have become the target of many drugs for cancer and infectious diseases [2, 16, 42, 70]. There are a variety of ways drugs can affect topoisomerases. Topoisomerase poisons interrupt the rejoining of the gate segment once the transported segment has passed through, resulting in damage to the DNA and possibly cell death [2, 70]. There has also been evidence that some anti-cancer agents can bind competitively to the ATP-binding sites of type II topoisomerase and thus inhibit its action [70]. Other drugs inhibit topoisomerase from binding to the DNA, preventing

multiple topoisomerase actions [2], or target DNA gyrase by inhibiting catalytic activity, such as supercoiling [70]. All these drugs prevent topoisomerases's normal actions, possibly leading to cell death. However, when topoisomerases fail to complete the unknotting and unlinking of DNA, a second type of enzyme, recombinase, has been found to have the ability to act in its place [5, 47, 53, 95]. If recombinase can act in topoisomerase's place *in vivo*, this has implications in drug treatments since targeting only topoisomerase will not be effective for treatments of cancer and infectious diseases.

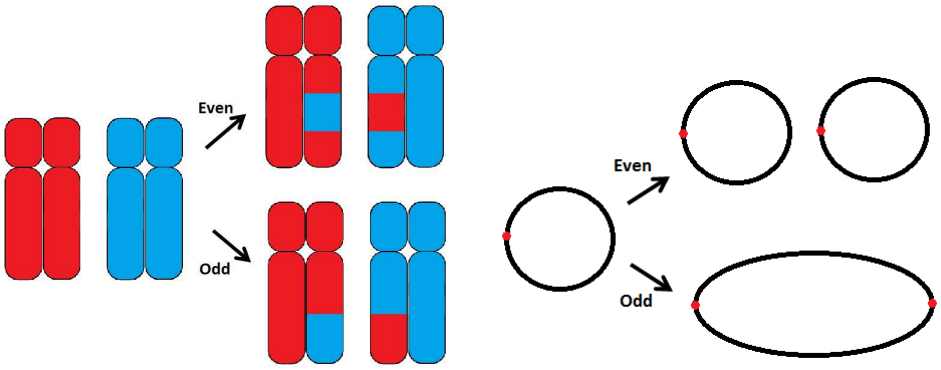
1.1.3 Recombinase Action on DNA

Recombination is "the rearrangement of a DNA molecule or molecules resulting from the breakage of the DNA duplex and its rejoining to another site" [12, page 120]. In this subsection the two main types of recombination, homologous recombination and site-specific recombination, will be introduced. The focus will be on site-specific recombination, specifically XerCD recombination which is responsible for the proper resolution of chromosome dimers in *E. coli*. It will be shown that this type of recombinase is capable of unknotting and unlinking DNA due to the position of its action and the presence of an additional enzyme, FtsK.

Homologous recombination takes place between two short homologous DNA sequences (sequences that are very similar), and is used for genetic diversity and to rebuild replication forks if they become broken or stalled [91]. In linear DNA this type of recombination is not a problem topologically, as an even or odd number of recombinations will result in DNA of the same form as the parent, *cf.* Figure 1.5 (a). However, if circular DNA becomes broken or stalled, homologous recombination enzymes act at the site to repair it. Occasionally errors in homologous recombination, or an odd number or occurrences, can yield a circular dimer, meaning both daughter chromosomes join into one circle, *cf.* Figure 1.5 (b), which was first noted by McClintock in 1932 [71]. In *E. coli*, on average once every six generations an error occurs during homologous recombinations and a circular dimer chromosome is generated [102]. Circular dimers are a problem for the cell, as daughter chromosomes joined together cannot properly segregate at cell division, and as a result the cell will die. A second type of recombination, site-specific recombination, is responsible for resolving these resulting chromosome dimers into two chromosome monomers.

Site-specific recombination (or simply recombination) is the rearranging of genetic sequences, and is a fundamental process of all chromosomes [92]. This type of recombination is called site-specific because the acting enzymes, recombinases, recognize and act at specific short sequences in DNA [12]. Site-specific recombination occurs between two sites and can result in either the integration, the excision, or the inversion of a segment of DNA, depending on the orientation of the sites [49], *cf.* Figure 1.6.

Site-specific recombination's mechanism can be broken into two general steps: first, the two specific sequences, called recombination sites, are brought together, forming a synapse, then the recombination enzymes break the two double-stranded DNA strands and exchange the ends in a specific manner [34, 49], *cf.* Figure 1.7 for one particular example. Because this operation exchanges DNA strands, it is called a *strand*



(a) (b)

Figure 1.5: Homologous recombination shown in linear (a) and circular (b) DNA. In the linear case, homologous recombination is not a problem topologically as the products that are the same form as the parent. However, in circular DNA errors in homologous recombination, or an odd number of occurrences, can result is the creation of a dimer, which is much different from the parent.

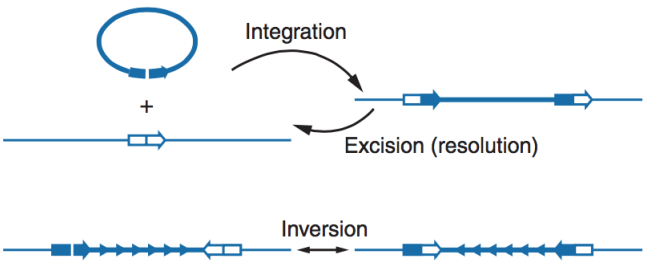


Figure 1.6: The three possible outcomes of site-specific recombination: integration, excision, or inversion. The selection of which occurs is based of the orientation of the sites being acted on. Reproduced with permission of Annual Reviews in the format Republish in a thesis/dissertation via Copyright Clearance Center [49].

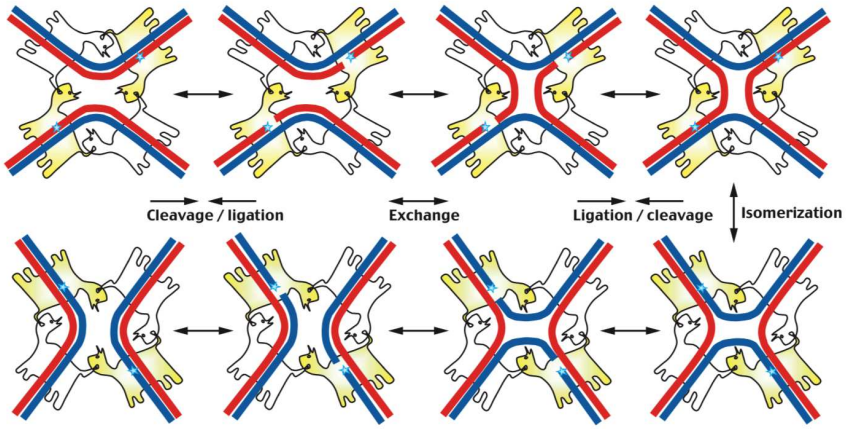


Figure 1.7: The mechanism for the tyrosine site-specific recombination family. Four recombination enzymes bind to the two double-stranded DNA sites shown in white and yellow. Two enzymes are initially activated (the initial yellow portions) and one strand is cut from each helix and exchanged (the red strands). The other two enzymes then become activated (the initial white portions) and the other strands are cut and exchanged (the blue strands). The rightmost panel is called a Holliday junction intermediate and is a characteristic of the tyrosine family. This operation exchanged the DNA strands, so is called a strand exchange. Reproduced with permission of Annual Reviews in the format Republish in a thesis/dissertation via Copyright Clearance Center [49].

exchange (SE). Afterwards, the synapse breaks apart, and the DNA product is released. There are two types of site-specific recombination families: serine and tyrosine. Both families follow these simple steps; however, their mechanisms for acting are different, and it is believed they evolved separately [49].

In the serine family, the two recombination sites are brought close together, then two double-strand breaks are made simultaneously at the recombination sites. Once all four strands are broken, a single 180° rotation of half the complex is made, before the ends are rejoined [49].

In the tyrosine family, the two recombination sites are again brought close together, but the two double-strand breaks occur separately. The tyrosine family is made up of many recombination enzymes which all use the same general idea for exchanging strands, shown in Figure 1.7. The two recombination sites are brought close together with four recombinase enzymes (the white and yellow portions of Figure 1.7) attached to the two strands of the double helices. Two enzymes are initially activated (the yellow portions of Figure 1.7) and one strand from each helix is cut and exchanged (the red strands of Figure 1.7). The other two enzymes then become active and the other two strands of DNA are cut and exchanged (blue strands of Figure 1.7) [49]. The four-armed intermediate in the rightmost panel of Figure 1.7 is called a Holliday junction and is a characteristic of the tyrosine family. By only breaking two strands at a time, as opposed to all four, the potential to incorrectly rejoin and form side products is reduced [49].

The tyrosine recombinases that are responsible for ensuring chromosome dimer resolution in *E. coli* are XerC and XerD [92]. The Xer recombination system consists of these two recombinases, XerC and XerD, at each recombination site separated by a central region [15, 49]. XerC, which binds to 11 bp, has been found to bind on the left side of the central region (as in the white portions of Figure 1.7), and XerD, which also binds to 11 bp, binds on the right side (as in the yellow portions of Figure 1.7) [15, 92]. Xer is the only system

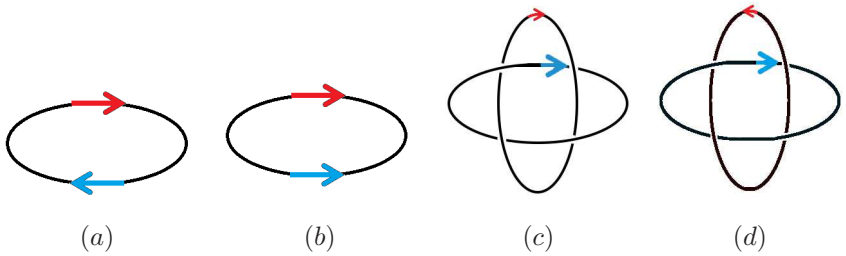


Figure 1.8: Possible orientations of recombination sites: (a) direct repeat, (b) inverted repeat, (c) parallel, or (d) antiparallel.

in the tyrosine family which uses two different recombinase enzymes, although XerC and XerD are similar with 37% identical amino-acid sequences [15, 23]. In Figure 1.7 the initial white pair of enzymes represents XerC and the initial yellow pair represents XerD, with the central regions between the enzymes. The use of two different enzymes appears to provide additional control over strand exchanges [8, 12]. The resolution of chromosome dimers by the Xer system is "considered one of the most conserved structural feature[s] of circular chromosomes in Bacteria and Archaea" [29, page 1], meaning Xer is used for chromosome dimer resolution in *E. coli* as well as many other systems. Modelling Xer recombinase will be the focus of this work.

As stated before, the result of site-specific recombination can be integration, excision or inversion, *cf.* Figure 1.6. The outcome depends on the linear orientation of the sequences of base pairs that constitute the recombination sites beforehand and whether the sites are in the same molecule or in separate molecules [34]. When both recombination sites are in the same molecule, the two sequence orientations relative to each other can be either *direct repeat* or *inverted repeat*, *cf.* Figure 1.8. *Direct repeat* means that each of the two sequences of the recombination sites appear on the same strand of DNA and in the same direction, for example $\frac{--CTTGA----CTTGA--}{--GAACT----GAACT--}$ (Figure 1.8 (a)), whereas *inverted repeat* means that each of the two sequences of the recombination sites appear on complementary DNA strands in opposite direction, for example $\frac{--CTTGA----TCAAG--}{--GAACT----AGTTC--}$ (Figure 1.8 (b)). When recombination sites are in different molecules with four or more crossings (see Section 1.2 for details), recombination sites are said to be *parallel* if both sequences are in the same order (Figure 1.8 (c)), and *anti-parallel* if not (Figure 1.8 (d)) [25]. When sites are in the same molecule, direct repeat causes deletion, creating two separate molecules (Figure 1.9 (a)), and inverted repeat causes inversion of the DNA segment between the sites (Figure 1.9 (b)). When sites are in different molecules, the result depends on the exact conformation of the molecules. Figure 1.9 (c) and (d) show that both parallel and antiparallel configurations change two molecules linked together into a single knotted or unknotted dimer. The result of site-specific recombination depends on the central region of the recombination site. Here, three different sites will be discussed: *cer*, *psi*, and *dif*. The *cer* and *psi* sites occur in *E. coli* plasmids (small genetic elements separate from chromosomal DNA in micro-organisms [36], such as *E. coli*) and require accessory proteins and sequences in order to occur. *Dif* sites occur in *E. coli* chromosomal DNA and are needed for the resolution of chromosomal dimers. The focus will be on *dif* as Xer acting at *dif* has been shown to both unknot and unlink DNA when topoIV is inhibited.

Cer and *psi* sites naturally occur in multicopy plasmids in *E. coli*, such as in ColE1 and pSC101 re-

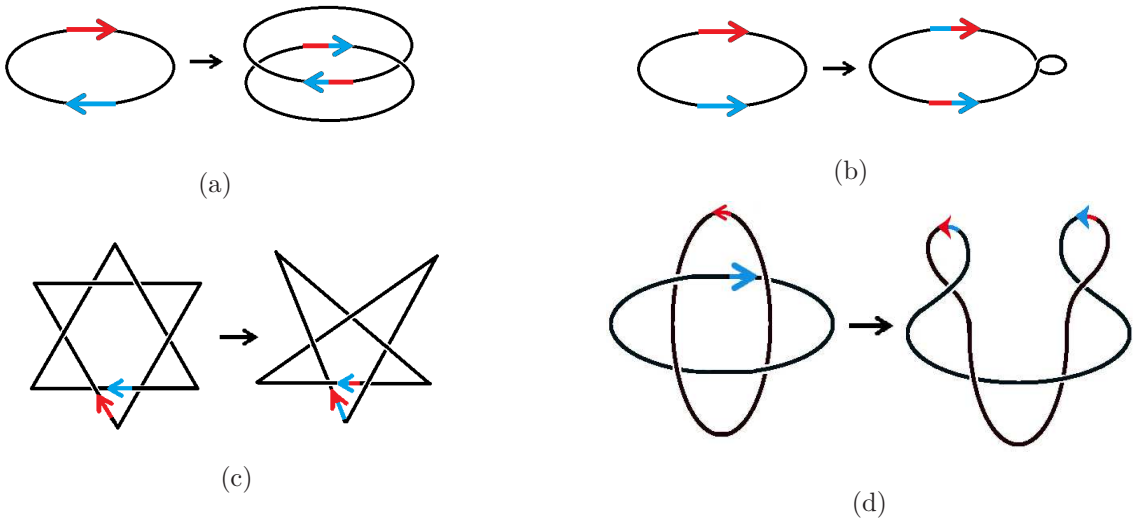


Figure 1.9: Result of recombination for: (a) direct repeats, (b) inverted repeats, (c) parallel sites, and (d) antiparallel sites.

spectively. Xer acts in plasmids in order to ensure multicopy plasmids, such as dimers, are resolved into monomers to avoid any segregation problems [22, 36]. These recombination sites not only require XerC, XerD, and the central *cer* or *psi* site, but also accessory proteins and sequences [92]. *Cer* is a central site of 8 bp and requires accessory proteins PepA and ArgR which bind to 178 accessory bp to the left of XerC [9, 22]. *Psi*, a 6 bp central site, also requires PepA, along with ArcA *in vivo*, which bind to 158 accessory bp [9]. A layout of these requirements can be seen in Figure 1.10.

For XerCD, to successfully perform a strand exchange at *cer* or *psi*, these accessory proteins and sequences are essential. This necessity has been shown to place a topological selectivity on XerCD recombination in plasmids, because a recombination event is only activated if the accessory sequences wrap around the accessory proteins trapping three right-handed crossings, *cf.* Figure 1.10 [22]. This selectivity ensures that XerCD acts at *cer* and *psi* only to resolve plasmid dimers to monomers, not to further complicate the monomers [22]. The particular synapse formed at *cer* in a dimer due to the accessory proteins and sequences is seen in Figure 1.11. It can be seen that the result of a strand exchange at this structure must be a four-crossing catenane which can be resolved by topoIV, therefore resolving the problematic dimer.

The third recombination site for XerCD in *E. coli*, *dif*, is an important site for recombination, since this is where chromosomal dimers are resolved, allowing replication and segregation to occur properly. *Dif* is 28 bp long, and is located in the termination region of the *E. coli* chromosome, meaning that it is one of the last parts of the chromosome to be replicated [22]. Early in experiments it was found that recombination only occurs at *dif* when chromosomal dimers are present, therefore ensuring that dimers are converted into chromosome monomers, not vice versa [102]. Initially, it was thought that this selectivity was because XerCD recombination at *dif* only occurred when replication was initiated [102]. However, it has since been found that replication is not necessary, but in order for recombination at *dif* to be successful in the cell, the protein

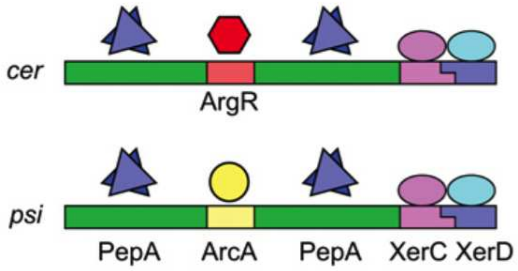


Figure 1.10: The central sites and accessory sequences and proteins for *cer* and *psi* Xer recombination. Reproduced from [22] in accordance with http://www.biochemsoctrans.org/content/how-do-i-find-out-about-rights-and-permissions.

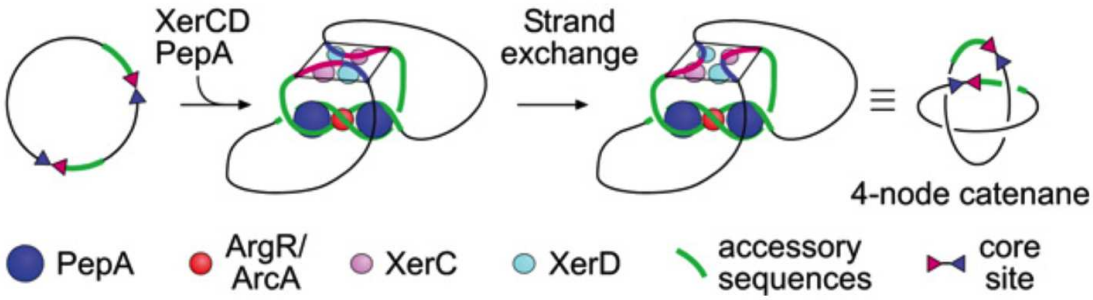


Figure 1.11: Example of Xer recombination at *cer* site with direct repeats, showing that a specific topology is required and a specific topology is the result. Reproduced from [22] in accordance with http://www.biochemsoctrans.org/content/how-do-i-find-out-about-rights-and-permissions.

FtsK must be present [5, 82, 101]. FtsK is a multifunctional protein that coordinates chromosome segregation and cell division [47], and directionally translocates on bacterial chromosomes, acting like a DNA pump [45]. FtsK is 1329 amino-acids in length and is broken into three domains *cf.* Figure 1.12: domain one, FtsK_N, localizes the protein to the mid-cell region; domain two, is the linker region and plays a part in cell division; and domain three, FtsK_C, acts as an orientated DNA pump towards *dif* sites and is involved in chromosome segregation [82]. It has been found that only FtsK_C is necessary for complete XerCD recombination and chromosome resolution *in vivo* [8]. Without FtsK, the pair of XerC enzymes are active and can initiate the first strand exchange, forming a Holliday junction, however, XerD is never activated. This means the second strand is not exchanged, so the resulting Holliday junction is resolved back to the starting state and no complete recombination occurs [8, 29, 49]. Grainge *et al.* [48] showed that if XerC and XerD are modified to be active without FtsK, recombination occurs but may produce complicated DNA knots and links, whereas when FtsK is present, recombination exclusively gives two unlinked monomers, *cf.* Figure 1.13.

Dif's location in the termination region of the cell plays an important role in recombination since this position allows XerCD at *dif* to differentiate between chromosome dimers and chromosome monomers, therefore ensuring that dimers are resolved and monomers are not made more complicated. The termination region in the cell is where replication finishes, and where chromosomal division takes place [82]. It has been seen that in order for *dif* to recombine efficiently to resolve chromosome dimers, it must be in a region 15-20 kilo bp

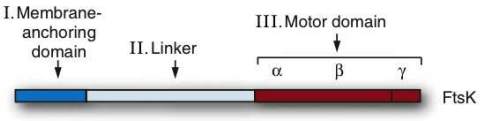


Figure 1.12: The domains of FtsK: I. $FtsK_N$, II. linker domain, and III. $FtsK_C$. [80] Reproduced with permission of NATURE PUBLISHING GROUP in the format Republish in a thesis/dissertation via Copyright Clearance Centre.

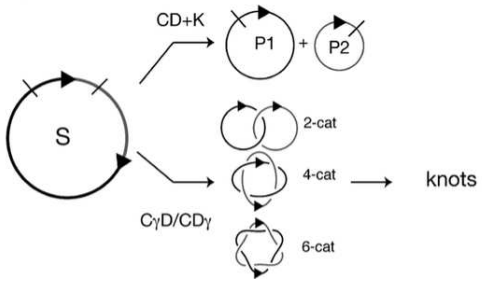


Figure 1.13: When XerCD recombination at two direct repeat *dif* sites occurs with FtsK (CD+K), the result is two free monomers, P_1 and P_2. Whereas, active XerC and XerD without FtsK ($C_\gamma D/CD_\gamma$) results in the formation knots and links. Reproduced from [48] under Creative Commons Attribution-NonCommercial 2.5 Generic License (CC BY_NC 2.5).

surrounding its normal position in the termination region [59, 77]. When *dif* is transplanted away from its normal position up to 13% of cell divisions fail, resulting in the death of the cell [78]. This reliance on the position of *dif* is due to the accessibility to FtsK.

As stated before, when FtsK (or a derivative with the third domain intact, such as $FtsK_C$ or $FtsK_{50C}$) is not present, XerC initiates a strand exchange but XerD is never activated, resulting in no recombination completion. However when FtsK is present, XerD is activated and forms the initial Holliday junction, which is then resolved by XerC both *in vitro* and *in vivo* [5]. It is thought that when FtsK is not present and XerC forms a Holliday junction which is not resolved by the inactive XerD, helps to stabilize the XerCD synapse. With this stabilization, FtsK can be loaded onto the DNA in order to translocate to the recombination synapse and activate XerD to resolve the chromosome dimer to monomers [5].

XerCD recombination at *dif* (XerCD-*dif*) takes place when a chromosome dimer exists due to homologous recombination, in order to resolve the dimer to monomers for proper segregation (*cf.* C. and D. in summary Figure 1.18 in Section 1.1.4). The segregation of chromosomes starts immediately after replication is started, so once the chromosomes have been replicated, they start to move towards the cell poles [93]. Therefore, if the chromosomes are monomeric, they are able to be immediately segregated away from mid-cell, whereas if a dimer is present it is unable to be segregated away, allowing FtsK access to the XerCD synapse and XerCD-*dif* recombination can take place [8]. So, with *dif* sites located mid-cell and FtsK localized here, XerCD is able to act to only resolve chromosome dimers to monomers, not vice versa. Figure 1.14 shows chromosome segregation involving XerCD, and Figure 1.15 shows a magnified view of the mid-cell region.

With FtsK present, XerCD-*dif* can act on supercoiled, relaxed or linear DNA, in either direct or inverted

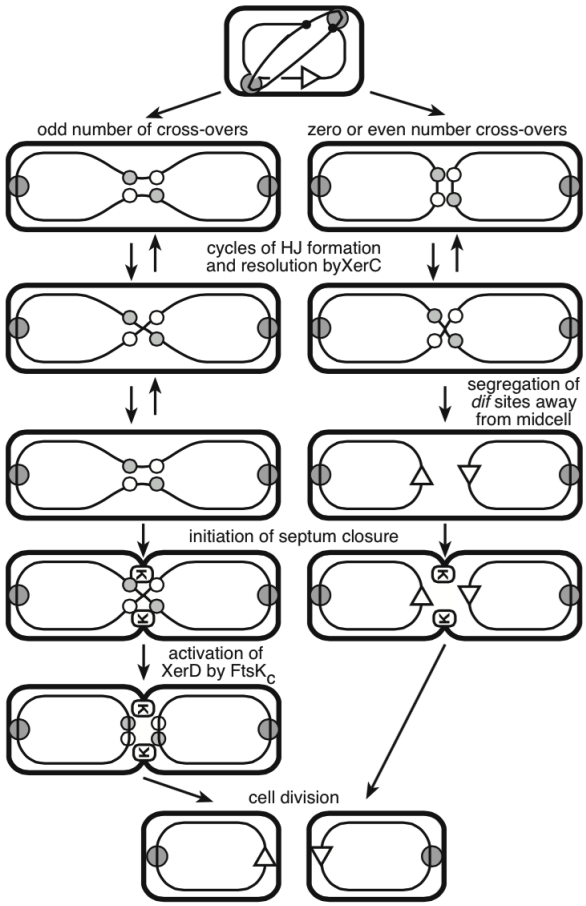


Figure 1.14: A display of chromosome segregation in *E. coli* which involves XerCD recombination. Replication origins (the large dark grey circles) are immediately segregated towards the cell poles, whereas the replication termination region, containing *dif* (the open triangle) stays mid-cell. Because the *dif* sites are kept mid-cell they can be synapsed by Xer recombinases (the small grey circle for XerC and the small white circles for XerD). If there are zero or an even number of homologous recombinations (the right side of the figure), XerC can form a Holliday junction which is cycled back until segregation breaks the synapse and forms the two daughter cells. If an odd number of homologous recombination events take place (the left side of the figure) a chromosome dimer is formed. When a dimer is formed, the synapse remains trapped mid-cell which allows FtsK (the rectangular K) to access the synapse and activate XerD. XerD forms the Holliday junction which is resolved by XerC, resulting in two chromosome which can be segregated successfully. Reproduced from [10] in accordance with http://www.pnas.org/site/aboutpnas/rightperm.xhtml.

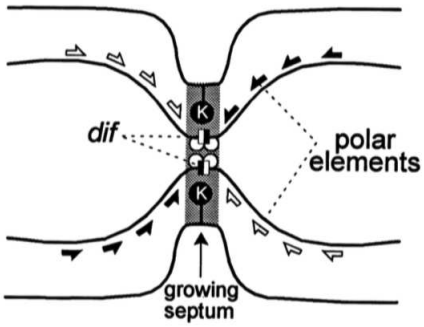


Figure 1.15: A magnified view of the mid-cell region of the cell during segregation. The black and white arrows represent orienting polar sequences which are orientated towards the *dif* sites to allow FtsK to access the *dif* sites when in this region. Here, *dif* sites are represented by a black and white squares, and XerC and XerD by white circles. The grey area represents the area *dif* must be in in order for FtsK to access it and activate XerD. If *dif* is outside this region successful dimer resolution cannot occur. Reproduced from [78] with permission from John Wiley and Sons Inc.

repeat orientation, showing the wide range of XerCD [5, 46, 53]. However, Ip *et al.* showed that $FtsK_{50C}$-dependent XerCD-*dif* is capable of not only converting dimer chromosomes to monomers, but also is capable of unlinking catenated DNA monomers with either antiparallel or parallel sites to free monomers *in vitro* [53]. They found that $FtsK_{50C}$-dependent XerCD-*dif* shows no discrimination between decatenation and dimer resolution so the unlinking process is possible through two steps. The first step is converting the catenane to an unknotted dimer, which is then resolved in the second step to two free monomers [53]. $FtsK_{50C}$ mediates the synapse formation ensuring that non-essential supercoils, catenations or knots do not get caught in the synapse, and also remodels the synapse to prepare for the next step. Therefore, XerCD-*dif* acts in a step-wise fashion with a reduction in link complexity at each step [53].

Grainge *et al.* then took the next step and were able to show that when topoIV is inhibited, XerCD-*dif* with FtsK (XerCD-*dif*-FtsK recombination) can act *in vitro* to unlink plasmid catenanes that accumulate in the absence of topoIV [47]. Figure 1.16 shows that, as Ip *et al.* [53] found, decatenation occurs in a stepwise manner [47]. It can be seen that over time, catenanes with up to 14 links are resolved to free monomer products within 30 minutes [47]. This shows that even if topoIV, which is essential for all cells, is blocked, XerCD-*dif*-FtsK can step in to decatenate DNA, allowing DNA to unknot and unlink, and the cell to divide and survive. Figure 1.17 shows how decatenation by topoIV compares to decatenation by XerCD-*dif*-FtsK. It should be noted that there can be many topoIV active at one time, while XerCD-*dif*-FtsK can only have one molecule active at a time, meaning topoIV is more efficient at decatenation [53]. Additionally, *dif* has been shown to be a preferred site for topoIV-DNA interactions, possibly because of the stable synapse formed by XerCD [8, 52], and it has also been shown that topoIV and $FtsK_C$ interact, with $FtsK_C$ stimulating topoIV's decatenation activity [35, 47].

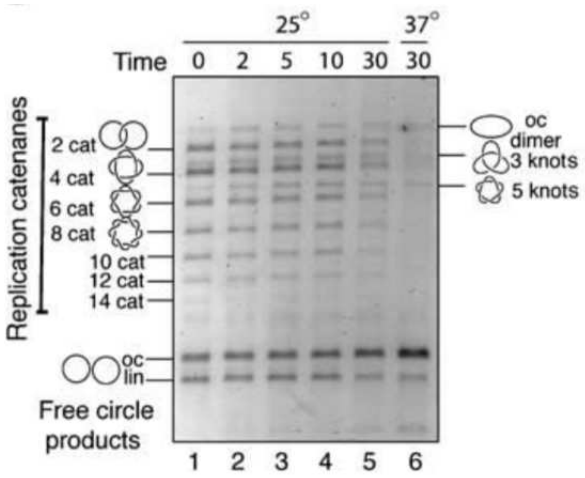


Figure 1.16: Gel electrophoresis showing that over time catenanes with up to 14 links (14 cat) produced *in vivo* can be unlinked with XerCD-dif-FtsK$_{50C}$ recombination *in vitro*. This unlinking process is step-wise with intermediate knotted and linked DNA steps. 'oc dimer' is an unknotted dimer, 'oc' are free monomer products and 'lin' are free for linear products. For details about gel electrophoresis see Section 1.2.3. Reproduced from [47] with permission from John Wiley and Sons Inc.

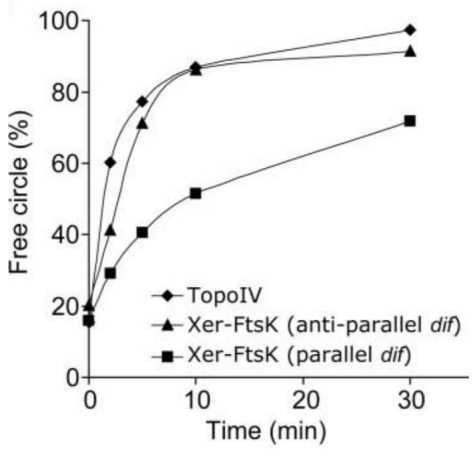


Figure 1.17: A comparison of decatenation by XerCD-dif-FtsK$_{50C}$ and topoIV taken at 20°C. It should be noted that there can be many topoIV active at one time, while XerCD-*dif*-FtsK can only have one molecule active at a time [53]. TopoIV is the most efficient at unlinking; however, XerCD-*dif*-FtsK can be comparable over time, especially when acting at anti-parallel site. Reprinted from [53] by permission from Macmillan Publishers Ltd.

1.1.4 Summary of Biology Background

A brief summary of the highlights of the biology background which are especially pertinent to this thesis will now be given.

It is known that topoIV is capable of relaxing DNA supercoils, and more importantly, due to its ability to break both strands of DNA, *in vitro* experiments have shown that topoIV is capable of unknotting and unlinking DNA [12, 16, 86]. Due to topoIV's ability to decatenate monomers and unknot dimers (A. and D. in Figure 1.18), it allows a cell to survive replication. Hence, it has become a target for drugs to treat cancer and infectious diseases [2, 16, 42, 70]. If topoIV is prevented from acting, it has recently been shown that recombinase has the ability to act in topoIV's place [5, 47, 53, 95].

The recombinase which is capable of this is Xer site-specific recombinase. Xer recombinase consists of two XerC and two XerD enzymes which bind to specific sites on the DNA called *dif* sites, *cf.* Figure 1.4. The bound strands are then cut and exchanged in a specific way, depending on the DNA sequence orientations of the *dif* sites, *cf.* Figure 1.9. XerCD-*dif* is capable of acting in a single DNA molecule, resulting in two separate molecules, or can act on two separate molecules, creating a new result, likely less knotted or linked than the original molecules.

It has been shown that XerCD-*dif* only performs a full recombination strand exchange *in vivo* when an additional enzyme, FtsK, is present [5, 82, 101]. FtsK is a multifunctional protein which is involved in chromosome segregation and cell division, and is naturally localized mid-cell [82]. Without FtsK, the XerC strand exchange occurs, but XerD is never activated, resulting in an incomplete process [8, 29, 49]. When FtsK is present XerD activates the first strand exchange and XerC is able to complete the recombination both *in vitro* and *in vivo* [5].

XerCD-*dif*-FtsK is able to differentiate between chromosome dimers and chromosome monomers, therefore ensuring to only act to resolve dimers and not to complicate monomers (C. and D. in Figure 1.18). This selectivity is possible due to *dif*'s position in the termination region of the chromosome. When replication takes place, the termination region is the last portion of the DNA to be replicated, meaning that it remains mid-cell the longest. As soon as replication begins, the new cells move towards the poles of the cell. If two monomers are present, replication will finish and the chromosomes will be segregated successfully. However, if a dimer is present, segregation cannot occur, meaning the termination region remains mid-cell [8]. As *dif* is mid-cell, the same region as FtsK, XerD is activated and recombination occurs, resolving the dimer (again, C. and D. in Figure 1.18), and allowing the resulting monomers to be successfully segregated.

Replication can result in dimers, but can also result in catenated monomers which are then also unable to be segregated properly. Ip *et al.* were able to show that XerCD-*dif*-FtsK is also capable of resolving this problem *in vitro* by unlinking catenated monomers with either parallel or antiparallel *dif* sites [53]. They showed that this unlinking process is a step-wise one, with the catenated monomers first being converted to a dimer and then two separated monomers [53]. Grainge *et al.* were then ultimately able to show that when topoIV is inhibited, recombinase can act *in vitro* to unlink cantenanes that have accumulated without

topoIV, also in a stepwise manner, *cf.* Figure 1.16 [47].

TopoIV is the main enzyme responsible for unknotting and unlinking DNA, and so is the most efficient; however, recombinase's unknotting and unlinking ability has been shown to be comparable to topoisomerase, *cf.* Figure 1.17, which has implications in drug treatments. If recombinase can act in topoisomerase's place *in vivo*, then only targeting topoisomerase will not be effective in drug treatments for cancer and infectious diseases.

The full DNA replication process with topoIV and Xer is shown in Figure 1.18. This figure does not show XerCD-*dif*-FtsK's ability to act to resolve DNA knots if topoIV is impaired (this would be at A., B. and E.).

It has been established that knots and links occur in DNA without a formal definition of knots or links. These definitions will be covered next.

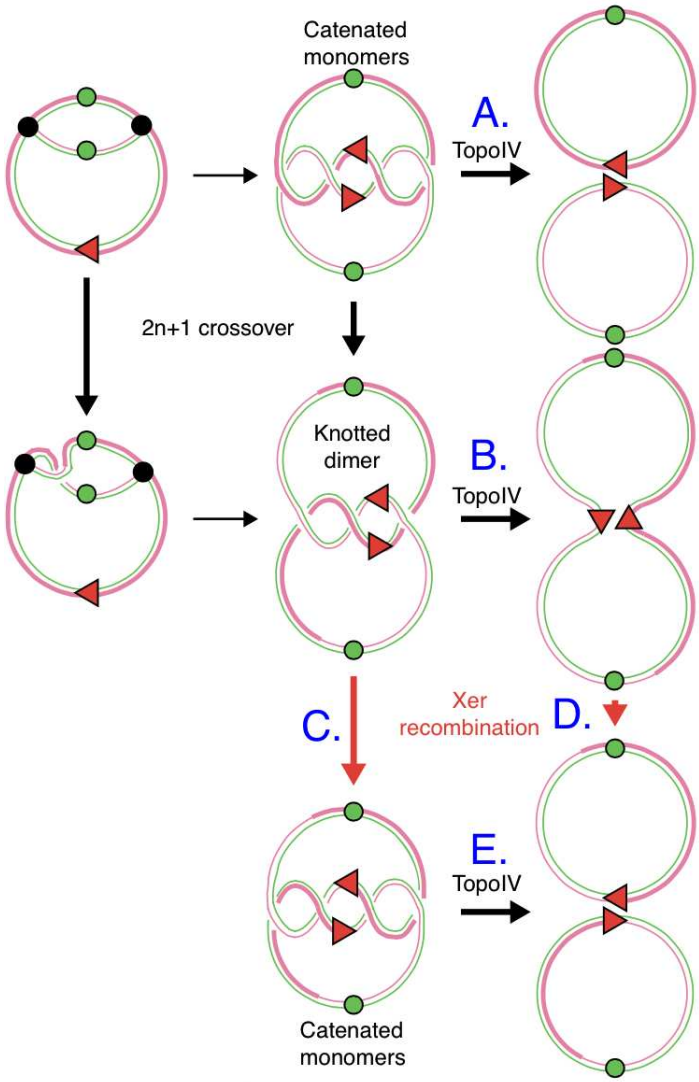


Figure 1.18: Replication of circular DNA. Green circles represent origin of replication, replication forks are black circles, and termination region is the red triangle. Replication starts at the origin and moves bidirectionally towards the termination region. If an error occurs during homologous recombinations, or an odd number or recombination events, the result will be a knotted dimer, otherwise catenated monomers are the result. If catenated monomers are generated then topoIV (A.) is able to resolve these to two separated monomers. If a knotted dimer is formed: topoIV (B.) can convert it to an unknotted dimer, which XerCD-dif-FtsK (D.) can then convert to two separate monomers; or XerCD-dif-FtsK (C.) can convert the knotted dimer to catenated monomers which topoIV (E.) can separate. Not included in this figure is the ability of XerCD-dif-FtsK to act to resolve knots if topoIV is impaired. Reprinted from [93] with permission from Elsevier.

1.2 Mathematical Background

In this section, the mathematical definitions of knots and links will be introduced and methods for identifying these computationally and biologically will be reviewed.

1.2.1 Knots

People often associate 'knots' with being found in shoelaces or causing problems in extension cords and headphones. However, these entanglement 'problems' are not knots as defined in mathematics. In mathematics, knots have a much stricter definition.

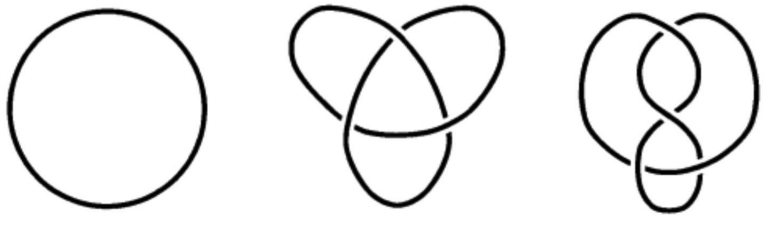
Definition 1.1 (Knot [24])**.** *Formally:*

$$A \text{ knot, } K \subset \mathbb{R}^3, \text{ is a subset of points homeomorphic to a circle.}$$

This means that a mathematical knot must be a closed curve in Euclidean three-dimensional space that does not intersect itself [1, 24]. Therefore, knots as defined above are not found in shoelaces, extension cords, or headphones because each of these are not closed curves.

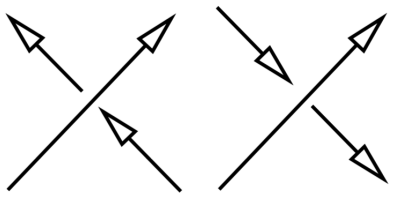
Two knots, K_1 and K_2, are said to be the *equivalent* if K_1 can be continuously deformed into K_2 without crossing itself [30]. All knots that are equivalent to each other form an *equivalence class*, referred to as the *knot type*, *i.e.* if K_1 can be deformed into K_2, the two knots have the same knot type. The circle has the simplest knot type and is called an unknot (Figure 1.19 (a)). If a knot is not equivalent to the unknot, then it is *knotted*. The simplest nontrivial knot with a regular projection having minimal crossing number three (see Section 1.2.3 for definitions) is the trefoil, (Figure 1.19 (b)), followed by the figure eight, with minimal crossing number four (Figure 1.19 (c)). Knot types are typically named in the form n_m, where n is the minimal crossing number and m represents the m-th knot type with minimal crossing number n as tabulated in [83] and [88]. For example, the unknot is denoted 0_1, or here simply ϕ; the trefoil is denoted 3_1; and the figure eight is denoted 4_1 [24].

If a knot K_1 is equivalent to its mirror image, then K_1 is *achiral*, which is the case for the figure eight. If a knot K_1 is not equivalent to its mirror image, then K_1 is called *chiral*. This is the case for the trefoil. To



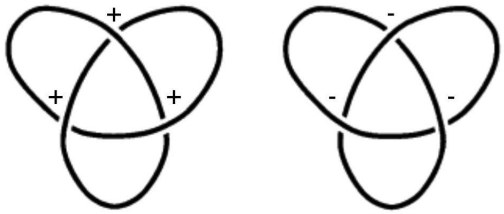
(a) (b) (c)

Figure 1.19: (a) An unknot, ϕ, (b) a trefoil, 3_1, and (c) a figure eight, 4_1. Each knot is seen with a regular project with minimal crossing number. Reproduced from [24] in accordance of http://www.cambridge.org/about-us/rights-permissions/permissions/permissions-requests/.



(a) (b)

Figure 1.20: (a) A positive crossing using the right-hand rule, (b) a negative crossing. Reproduced from [106], edited by Cheston.



(a) (b)

Figure 1.21: (a) A positive trefoil, 3_1^+, (b) a negative trefoil, 3_1^-. Reproduced from [68, 67] under Public Domain, edited by Cheston.

distinguish between two knots in a chiral pair (a chiral knot and its mirror image), knots can be labelled as either positive or negative (as proposed by [79]). A positive crossing (Figure 1.20 (a)) follows the conventional right-hand rule, while a negative crossing (Figure 1.20 (b)) does not. For the trefoil, in one case all three crossings are positive when using the right-hand rule, so the knot is called 3_1^+ (Figure 1.21 (b)), and in the other case all the crossings are negative, so called 3_1^- (Figure 1.21 (c)). If the figure eight was mirrored and this crossing labelling convention was applied, its achiral property would be seen as there would be only one unique labelling when deformation is taken into account.

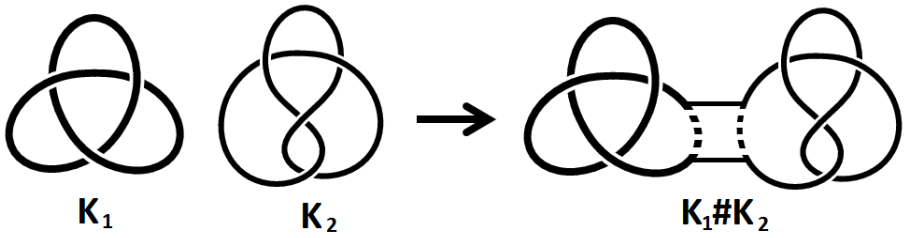
If a knot is made of two or more independent knots joined together such as in Figure 1.22, then the resulting knot is called a composition of the two knots [1]. A prime knot is one that cannot be decomposed into two or more prime knots. All knots discussed thus far have been prime knots. A knot that is composed of two or more prime knots is called a *composite knot* [1]. The composition of K_1 and K_2 is represented by $K_1 \# K_2$, and an example of a trefoil composed with a figure eight can be seen in Figure 1.22 (a). Any knot composed with the unknot, remains itself, as seen in Figure 1.22 (b).

1.2.2 Links

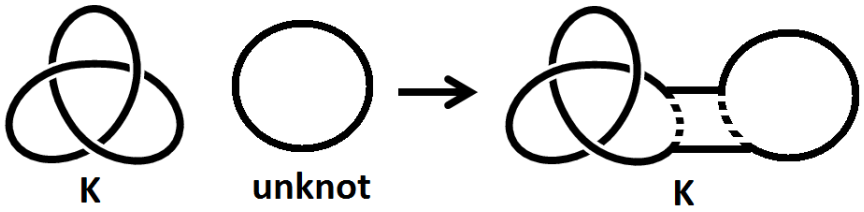
So far, we have talked about having single knotted closed curves. It is possible to have several closed curves embedded in three-space where each curve is a knot; the resulting object is called a link.

Definition 1.2 (Link [24])**.** *A* link *is a finite union of knots:*

$$L = K_1 \cup K_2 \cup \cdots \cup K_n.$$



(a)



(b)

Figure 1.22: (a) A composite knot of 4_1 and 3_1 creating $4_1\#3_1$ (b) a composite knot of 3_1 and ϕ creating 3_1. Reproduced from [66, 67] under Public Domain, edited by Cheston.

(a) (b)

Figure 1.23: (a) A trivial link, the unlink and (b) a Hopf link, 2_1^2.

Like with knots, two links, L_1 and L_2, are said to be equivalent if L_1 can be continuously deformed into L_2. Again, all links that are equivalent to each other form an equivalence class known as a *link type*. The simplest link consists of only one component, which is a knot; therefore knots are a subset of links. In this work 'link' will also include single component links, *i.e.* knots. The simplest multi-component link is the trivial link, or an unlink, which is multiple disjoint unknots (Figure 1.23 (a)). If a two-component link is not the unlink, then it is *linked*. The simplest multi-component, nontrivial link is called the Hopf link, which is two unknots linked together (Figure 1.23 (b)). Link types are typically named in the form n_m^P, where n is the minimal number of crossings in the link, P is the number of components of the link, and m represents the m-th link with n crossings and P components as tabulated in [83] and [88]. Therefore, the unlink is 0_1^2 and the Hopf link is 2_1^2.

Also like knots, links can differ in their crossing signs depending on the orientation of the links, affecting their chirality. The Hopf link is a chiral link [24]. Each component of a link has its own orientation which then gives a linked orientation between the components. Figure 1.24 shows the two possible orientations for the Hopf link, which in this work will be considered different link types.

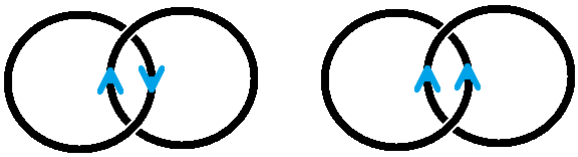


Figure 1.24: The two possible orientations of the Hopf link, 2_1^2. In this work they are considered different link types.

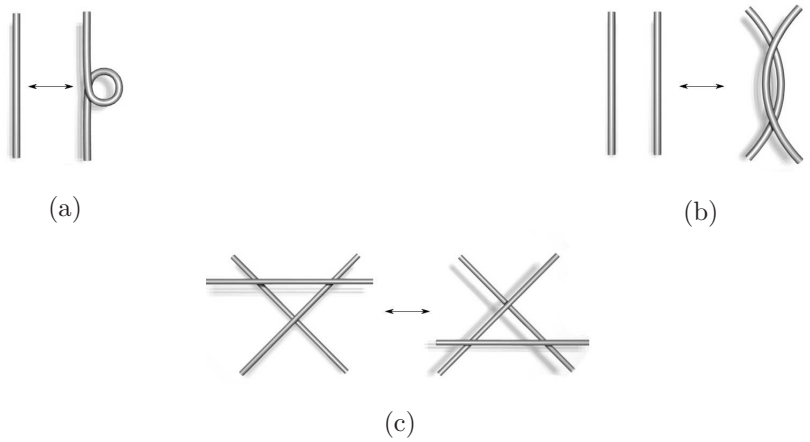


Figure 1.25: Reidemeister moves: (a) type I (b) type II (c) type III. Reproduced from [118, 119, 120] under Creative Commons Attribution-ShareAlike 3.0.

1.2.3 Identifying Knots and Links

In practice it is not easy to identify whether two knots (or links) are equivalent, so a knot projection is often used. A projection of a knot is obtained when a knot in $\mathbb{R}^3$ is projected into $\mathbb{R}^2$, with positive and negative crossing information retained. A projection is said to be *regular* if there are only a finite number of double points with no other type of multiple points allowed [76]. Two regular knot projections (and so the knots) are equivalent if one projection can be deformed into the second projection through a finite series of specific moves called Reidemeister moves [1]. There are three different Reidemeister moves: a type I Reidemeister move either puts in or takes out a twist in the knot, *cf.* Figure 1.25 (a); a type II move can either add or remove two crossings, *cf.* Figure 1.25 (b); and a type III Reidemeister move allows a strand of the knot to slide from one side of a crossing to the other side, *cf.* Figure 1.25 (c). So, knot K_1 is equivalent to knot K_2 if a series of Reidemeister moves transforms a knot projection of K_1 into a knot projection of K_2.

The *minimal crossing number* of a knot K is the minimal number of crossings over all regular knot projections equivalent to K.

Computationally, there exist a number of algorithms that identify the knot type of a knot. The one used in this work is the HOMFLY polynomial, which is a generalized version of two previous polynomials, the Alexander polynomial and the Jones polynomial [41]. Each knot type has a corresponding HOMFLY

(Hoste, Ocneanu, Millett, Freyd, Lickorish and Yetter) polynomial, so by calculating the polynomial for a specific knot, the knot type can be determined. In this work the HOMFLY polynomial calculations are performed using KnotPlot [88]; the details of this algorithm are beyond the scope of this work, and hence are not given. It should be noted that HOMFLY polynomials are not necessarily unique; Jones [54] showed that the HOMFLY polynomial is not unique for every knot type, for example 5_1 and 10_{132} have the same HOMFLY polynomial. Additionally, the polynomial has trouble differentiating chirality for some knot types; for example, the chiral pair of knots associated with 9_{42} give the same HOMFLY polynomial. In this work the majority of knots considered have a low minimal crossing number, so these limitations will not play a significant role.

Identification of knots using polynomials is very useful mathematically; however, these methods cannot be directly applied biologically to identify DNA knots. Biologically, there are two main ways in which DNA knots are observed and identified [16]. The first method is via agarose gel electrophoresis, which provides a rapid and quantitative measure of knotting [27]. As per Buck [16], to perform agarose gel electrophoresis the backbone of the DNA must first be nicked in order to release any supercoils. Then, an agarose gel slab (a polymer material generally extracted from seaweed) is placed in an aqueous bath with the DNA samples in one end of the gel. A current is run through the bath, and because DNA has a negative charge, the DNA will migrate to the positive end of the gel. The gel electrophoresis will then separate the DNA into knots and links by minimal crossing number. Figure 1.16 shows a picture of the result of gel electrophoresis as seen under UV light. The different bands correspond to different knot and link types that migrate through the gel at different speeds depending on their complexity. Knots with greater complexity, *i.e.* more crossings, migrate more quickly than those with few crossings [16]. It has been found that there is a linear relationship between crossing number and speed [100]. However, a limitation to agarose gel electrophoresis is that knots with the same crossing numbers cannot be distinguished; for example, 5_1 and 5_2 would not be able to be distinguished.

Due to this gel electrophoresis limitation, it is often used along with the second method of identification, electron microscopy. Electron microscopy is how Liu *et al.* first observed double stranded DNA knots since it allowed the DNA to be easily visualized [16, 60]. The DNA molecule is first coated with RecA, a protein which stiffens and thickens the molecule. This coating allows the precise knot or link to be identified, since the thickening of the molecule enables each crossing to be identified as a positive or negative crossing, *cf.* Figure 1.26 [16]. The knot type can then be identified via the resulting projection. It should be noted that this process is very difficult and expensive to perform.

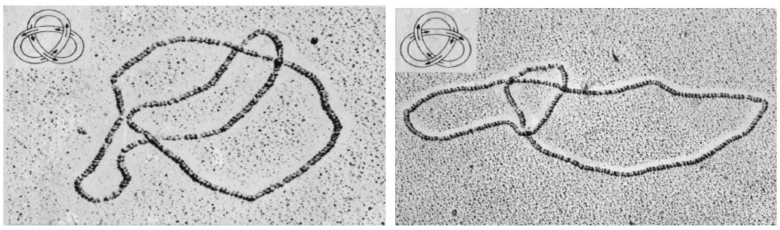


Figure 1.26: Two electron microscopy pictures of recA coated DNA, a 3_1^+ and a 3_1^-. Reprinted from [58] by permission from Macmillan Publishers Ltd.

1.3 Enzyme-DNA Models Review and Questions for LSE Model

A brief review of previous models of enzyme-DNA interactions will now be provided.

In Section 1.1 it was stated that polymers can be used to model macroscopic features of circular DNA by the axis of the DNA representing the polymer. One standard method for modelling ring polymers is using polygons [24, 114]. A polygon is "a finite set of straight line segments in $\mathbb{R}^3$ which intersect only at their endpoints" [24, page 5]. The straight line segments are edges which end at vertices, so exactly two edges meet at one vertex. Therefore, DNA is modelled by a polymer, which is then modelled with a polygon; this progression can be seen in Figure 1.27.

As reviewed earlier in Section 1.1.2, topoisomerase-DNA interactions have been of interest due to topoisomerase's efficiency at unknotting and unlinking DNA. DNA experiments have shown that topoisomerase is much more efficient at unknotting and unlinking DNA than if it was acting at random locations within the DNA [86]; however determining the rules that govern where topoisomerase binds to DNA is an open problem. Several random polygon models have been used to address this problem [17, 20, 38, 62, 64, 89, 105, 106, 107, 108]. Some of these models use worm-like chains which are off-lattice models capable of incorporating bending and torsional energy, so are considered more DNA-like; while others use lattice models which are a coarser approximation to DNA. One goal of these models is to investigate how, by acting only locally (via a strand passage), a global property such as knot type can be efficiently simplified. This has been investigated by exploring the dependence of knot and link transitions and the knot reduction factor on local information at the strand passage site.

Specifically, knot transition probabilities, $\rho_n^{SP}(K \to K')$, measure how a single modelled strand passage

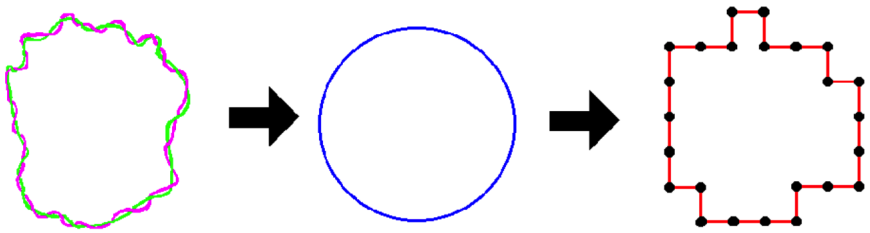


Figure 1.27: Circular DNA can be modelled as a polymer which can be modelled as a polygon. Reproduced from [33] with permission from author.

can change the knot type of the polygon, where n is the number of edges of the polygon, SP indicates the strand passage model, K is the knot type before the strand passage and K' is the knot type after the strand passage. These probabilities have been looked at for multiple models. The knot reduction factor measures how much the unknot/knot population ratio is changed by a modelled strand passage action [86, 107], which has been used to compare to DNA experiments. Additionally, information concerning the local geometry of the two juxtaposed strands at the strand passage site has been of interest, to investigate if different strand geometries affect the transition probabilities [107]. For a lattice model, knot transition probabilities have been shown to match conjectured polymer scaling theory as polygon length goes to infinity [20]. Many of these studies found that a hook-like geometry leads to greater knot and link reduction than other geometries, and recently, using the LSP, it was shown that knot reduction also depends on the crossing sign change of the strand passage [107].

There have also been models for modelling recombinase-DNA interactions, mostly focusing on visualization and using a method called tangle analysis [16, 26, 87, 95, 115]. Tangle analysis uses mathematical objects called *tangles* to model the changes to the recombinase synapse during recombination by comparing the synapse before and after a recombination event [95]. The tangle method relies on only a few assumptions, and can produce a set of tangle equations, which represent the product of recombination. Tangle analysis has been used to mathematically prove multiple biological results. As stated in Section 1.1, when XerCD recombination occurs at a *psi* site, the result will be a four-crossing link, *cf.* Figure 1.11 [22]. With tangle analysis it was proved that there are only three tangle solutions consistent with experimental data and that all three solutions are different projections of the same three-dimensional object [115]. Therefore, Vazquez *et al.* were able to propose a unique topological mechanism of XerCD acting at *psi* to incorporate all three solutions [115]. Furthermore, tangle analysis has proven that the shortest path for unlinking a catenane with $2m$-links with XerCD-*dif*-FtsK is exactly $2m$ steps, and this path is unique, *cf.* Figure 1.28 [95]. This means that tangle analysis reinforces the idea that XerCD-*dif*-FtsK acts iteratively, reducing the complexity of the link in each step.

The information that recombinase is able to unlink and unknot DNA is relatively new, so very little work has been done to investigate transition probabilities associated with recombinases' action. In this thesis a new simple cubic lattice model, the LSE model, is introduced in order to model the recombinase-DNA interaction and to investigate link transition probabilities after a recombinase strand exchange has occurred. These link transition probabilities, $\rho_n^{SE}(K \to L)$, are the probability of going from knot type K to link type L after a strand exchange occurs in an n-edge polygon, accounting for both direct repeat sites and inverted repeat sites. This analysis can then be compared to biological results and to the previous tangle method of modelling recombinase in order to investigate whether the same unlinking property is seen in the LSE model. Specifically, in Chapter 4 the asymptotic properties of $\rho_n^{SE}(K \to L)$ as $n \to \infty$ (also for each strand exchange operation) will be investigated and, as in [20], the asymptotic form will be verified to be consistent with one obtained from polymer scaling theory. The LSE model will also be compared to the

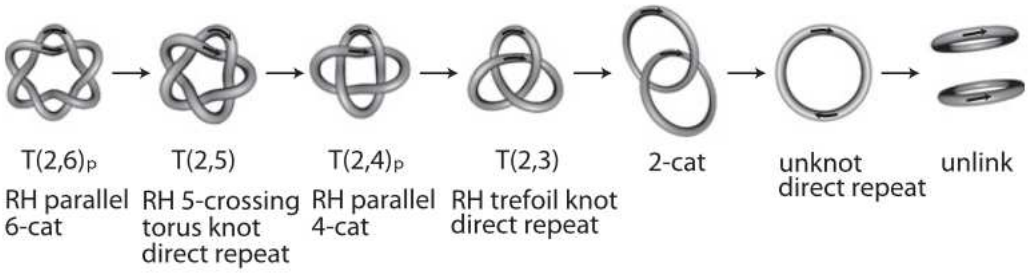


Figure 1.28: The unlinking path of a right-handed (positive) $2m$-catanene to two separate circular products. A step-wise process where the complexity is reduced with each step. Reproduced from [95] in accordance with http://www.pnas.org/site/aboutpnas/rightperm.xhtml.

similar topoisomerase model, LSP, to see whether trends exist across the models. In Chapter 5 polygons with similar SAW lengths will be investigated in order to model *dif*'s position in the termination region of the cell (located mid-cell), resulting in a dimer splitting into two equal halves. Transition probabilities will be investigated for this specialized subset in order to compare to XerCD-*dif*-FtsK's action. Indeed evidence is provided that, in this biologically relevant situation, topology simplification is the most probable path, consistent with experiments. The LSE model is the first-step in moving towards a more realistic model of the recombinase-DNA interaction, while also creating a model for general enzyme-DNA interactions.

1.4 Thesis Outline

The remainder of the thesis is outlined as follows. Chapter 2 will begin with an explanation for how DNA can be modeled with self-avoiding polygons (SAPs), and then provide the introduction of statistical definitions, particularly the link transition probability, which will be a main focus of this work. The chapter will then include a review of the Local Strand Passage Model (LSP) created by Szafron and Soteros [105, 106, 107, 108] for modelling topoisomerases's interaction with DNA. Finally, Chapter 2 will describe the newly created Local Strand Exchange Model (LSE) to model recombinases's interaction with DNA. Chapter 3 will focus on the statistical theory and algorithms used to generate the SAP data and on the analysis techniques used to analyze this generated data. Much of this data generation and analysis relies on previous work for the LSP model, this work will be briefly reviewed in Chapter 3 for use with the LSE model. Chapter 4 will present the results for the LSE model. First, the analysis techniques outlined in Chapter 3 will be reviewed for the particular LSE simulation, and then link transition probabilities for the LSE model will be presented. The chapter will conclude with a comparison of the LSE model to the LSP model. Chapter 5 will focus on a specialized subset of SAPs which will be compared to biologically and mathematically motivated results. Finally, Chapter 6 will give conclusions and describe potential future work. Appendix A contains additional data analysis from Chapter 4; Appendix B contains tables for all observed link transitions in the LSE model; and Appendix C contains data relating the LSP and LSE models.

Chapter 2

Model

Polymers (and DNA) almost always exist in solution, so they are immersed in a solvent [114]. A good solvent is one where it is energetically favourable for the monomers of the polymer to be surrounded by the solvent rather than by other monomers. This means that there exists a small volume surrounding a particular monomer where there is a low chance of finding another monomer. This property, called the excluded volume property, leads to a more open, expanded structure for the polymer [114]. This property is important to consider when modelling DNA. One of the most popular polymer models which takes this property into account is the self-avoiding polygon (SAP) model, in which polymer (such as DNA) configurations are represented by SAPs on the simple cubic lattice.

This chapter will introduce the self-avoiding polygon models which are used to model DNA in this thesis. The model for topoisomerase's action, the Local Strand Passage model, will then be reviewed, and then the model for recombinase's action, the Local Strand Exchange model, will be introduced. Since we are interested in modelling DNA, which is very long, we are interested in polygons which are also long. For this reason we are also interested in properties of polygons when polygon length goes to infinity, called asymptotic properties. Therefore, throughout the chapter necessary terminology and statistical background will be introduced in order to investigate polygon properties, namely the link transition probabilities, for varying lengths and as the lengths of the polygons go to infinity.

2.1 Self-Avoiding Polygons on the Simple Cubic Lattice

In this section self-avoiding polygons (SAPs) on the simple cubic lattice will be introduced as SAPs will be used to model DNA. Since we are ultimately interested in link transition probabilities as polygon lengths go to infinity, a number of model and statistical definitions are introduced next.

Definition 2.1 (Simple Cubic Lattice [4])**.** *The* simple cubic lattice *is defined to be the infinite graph embedded in* $\mathbb{R}^3$ *with vertex set* $\mathbb{Z}^3$ *and edge set* $\{\{u, v\} | u, v \in \mathbb{Z}^3, |u - v| = 1\}$*, where* $|u - v|$ *is the Euclidean distance between* u *and* v*.*

That is, each vertex of the lattice is represented by a 3-tuple, (x, y, z), of integer values from $\mathbb{R}^3$.

Definition 2.2 (Self-Avoiding Walk (SAW) [98])**.** *An* n*-edge* self-avoiding walk (SAW) *starting at the origin on the simple cubic lattice is an alternating sequence of* $n+1$ *distinct vertices and* n *directed edges:*

$u_0, (u_0, u_1), u_1, (u_1, u_2), u_2, \cdots, u_{n-1}, (u_{n-1}, u_n)$, *such that the vertices* ($u_i \in \mathbb{Z}^3$ *for* $i = 0, \ldots, n$), ($u_0 = (0, 0, 0)$), *and for each* $i = 0, \ldots, n-1$ *the directed edge* (u_i, u_{i+1}) *joins two nearest neighbour vertices, i.e.* $|u_{i+1} - u_i| = 1$. *This means that the walk never repeats a vertex or an edge, and two vertices* u_i *and* u_{i+1} *only ever differ in one coordinate, with the difference being* $\pm$ *1.*

An n-edge SAW is known to be a good model for a linear polymer with n monomers, in a good solvent [114]. In order to study circular DNA, we need circular polymers which can be modelled with self-avoiding polygons [114].

Definition 2.3 (Self-Avoiding Polygon (SAP) [98])**.** *An* n*-edge* self-avoiding polygon (SAP) *in* $\mathbb{Z}^3$ *is an alternating sequence of* n *distinct vertices and* n *distinct undirected edges:* $u_0, \{u_0, u_1\}, u_1, \{u_1, u_2\}, u_2, \cdots, u_{n-1}, \{u_{n-1}, u_0\}, u_0$, *such that for each* $i = 0, \ldots, n-1$ *the vertex* $u_i \in \mathbb{Z}^3$, *and the edge* $\{u_i, u_{i+1}\}$ *joins two nearest neighbour vertices, i.e.* $|u_{i+1} - u_i| = 1$.

An n-edge SAP can be viewed as an $(n-1)$-step SAW where the $(n-1)$-st step of the walk brings it back to a nearest neighbour site of the starting point and where the directions on the walk edges are ignored [114]. With SAWs and SAPs defined, a number of properties for each will now be defined.

Definition 2.4 (Rooted Polygon and Unrooted Polygon [106])**.** *For a SAP,* $\omega \in \mathbb{Z}^3$, ω *is referred to as a* rooted polygon *if one of its vertices is designated as the root of* ω. *If no vertex is specified, then* ω *is an* unrooted polygon.

Definition 2.5 ($|\omega|$ [65])**.** *For any SAW or SAP,* ω, *the length of* ω, $|\omega|$, *is the number of edges in* ω.

Definition 2.6 (c_n and p_n [65])**.** *Define* c_n *to be the number of distinct* n*-edge SAWs and* p_n *to be the number of distinct* n*-edge SAPs, where SAPs are said to be distinct if one SAP cannot be obtained from another by a lattice translation.*

Since we are interested in looking at the knot type of SAPs, further definitions are useful.

Definition 2.7 ($k(\omega)$ [106])**.** *For any SAP* ω, *define* $k(\omega)$ *to be the knot type of* ω.

Definition 2.8 ($p_n(K)$ [106])**.** *Define* $p_n(K)$ *to be the number of distinct* n*-edge SAPs with knot type* K.

Definition 2.9 ($\rho_n(K) \coloneqq \frac{p_n(K)}{p_n}$ [20])**.** *The probability that an* n*-edge SAP has a particular knot type,* K, *is given by* $\rho_n(K) \coloneqq \frac{p_n(K)}{p_n}$, *i.e. the number of* n*-edge polygons with knot type* K *(up to translation) divided by the total number of* n*-edge polygons (up to translation).*

There are a number of benefits to the self-avoiding polygon model on the simple cubic lattice. As stated earlier, this model takes the excluded volume property into account which is important when modelling DNA. Additionally, it is expected that lattice models, such as this one, exhibit some behaviours that are the same as 'real' polymers; *e.g.* "field theoretic arguments suggest that there exist universal quantities, *i.e.* critical exponents, which will be exactly the same for lattice models and for real polymers" [98, page 3]. By moving

from $\mathbb{R}^3$ to $\mathbb{Z}^3$ through the use of lattice polygon models, conformational freedom is not significantly reduced and rigorous mathematical analysis, such as combinatorial and asymptotic analysis, is possible [98].

Although c_n and p_n can be defined, determining them is computationally difficult. The highest enumerations achieved are $p_{32} = 53,424,552,150,523,386$ by Clisby *et al.* in 2007 [21] and $c_{36} = 2,941,370,856,334,701,726,560,670$ by Schram *et al.* in 2011 [90]. The smallest knotted polygon, the trefoil, has 24-edges, and the figure eight has 30-edges, which is followed by 5_1 with 34-edges [110]. So by enumeration, only the trefoil and figure eight knots would be able to be investigated. This difficulty is a problem as DNA is typically very long, so polygons modelling DNA must also be long. Therefore, we are also interested in p_n as n goes to infinity. In 1954, Hammersley and Morton determined that c_n grows exponentially with n:

Theorem 1 (Hammersley and Morton [51]). *The following limit exits:*

$$\kappa = \lim_{n\to\infty} \frac{1}{n} \log c_n.$$

where κ is referred to as the connective constant *for SAWs in $\mathbb{Z}^3$.*

Definition 2.10 ($\mu := e^{\kappa}$ [106]). *The* growth constant *for SAWs in $\mathbb{Z}^3$ is defined to be $\mu := e^{\kappa}$.*

From here, Hammersley proved that the connective constant, and so the growth constant, for SAPs was equal to that of SAWs,

Theorem 2 (Hammersley [50]). *The following limit exists:*

$$\lim_{n\to\infty} \frac{1}{n} \log p_n = \kappa.$$

Hence, as $n \to \infty$ both c_n and p_n grow at the same exponential rate. Sumners and Whittington went on to prove the following for the unknot:

Theorem 3 (Sumners and Whittington[103]). *The following limit exits:*

$$\kappa > \kappa_\phi := \lim_{n\to\infty} \frac{1}{n} \log p_n(\phi).$$

And so by Definition 2.10 and Theorem 2:

Definition 2.11 (μ_ϕ [106]). *The growth constant for unknotted SAPs is $\mu_\phi := e^{\kappa_\phi}$.*

Theorem 3 states that the connective constant, κ_ϕ, and its associated growth constant, are less for unknotted SAPs than for all SAPs, *i.e.* $\kappa_\phi < \kappa$, which means:

$$\text{as } n \to \infty : \frac{p_n(\phi)}{p_n} = \rho_n(\phi) \to 0. \tag{2.1}$$

Hence the probability of a SAP being unknotted goes to zero as n goes to infinity. Looking at this in terms of the probability of a SAP being knotted (*i.e.* any knot that is not the unknot), gives:

$$\text{as } n \to \infty : 1 - \frac{p_n(\phi)}{p_n} = 1 - \rho_n(\phi) \to 1. \tag{2.2}$$

Soteros *et al.* [99] then showed that for any particular knot type $K \neq \phi, \kappa_K \coloneqq \limsup_{n\to\infty} \frac{1}{n} \log p_n(K) \leq \kappa_\phi < \kappa$, meaning that the probability of a SAP being a specific knot type K also goes to zero:

$$\text{as } n \to \infty : \frac{p_n(K)}{p_n} = \rho_n(K) \to 0. \tag{2.3}$$

Based on polymer scaling theory arguments, and in analogy with the expected asymptotic form of p_n, Orlandini *et al.* [75] conjectured the asymptotic form for $p_n(K)$ as $n \to \infty$:

$$p_n(K) \sim A_K n^{\alpha_0 + f_K} \mu_0^n \tag{2.4}$$

for the number of n-edge polygons with fixed knot type K, where α_0 and μ_0 are constants and f_K is the number of prime knots in the prime-knot decomposition of K (note, for $K = \phi$ it is assumed that $f_K = 0$). It is expected that μ_0 is model dependent, however α_0 is thought to be universal, *i.e.* does not depend on the model being used. This conjecture has been supported by numerical analysis from multiple groups [6, 7, 20, 69, 75, 81, 110, 111, 112].

The SAP model for topoisomerase's action, the Local Strand Passage (LSP) model, will be reviewed next.

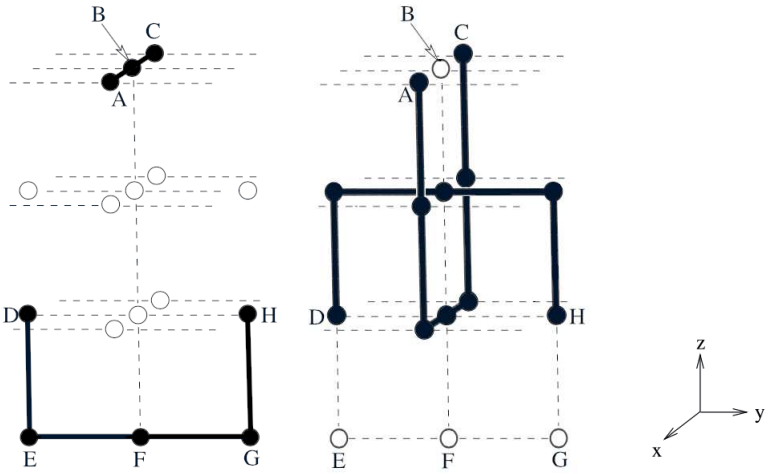
2.2 The Local Strand Passage (LSP) Model

Section 2.1 introduced SAPs, and showed we can model polymers, and so DNA, using SAPs. In 2000, Szafron proposed a model called the Local Strand Passage (LSP) model in order to model type II topoisomerase action on DNA using SAPs [105, 106]. This model will be reviewed in this section, along with conjectures and results about the associated knot transition probabilities as the lengths of the SAPs go to infinity.

The LSP model uses a fixed region in the polygon, denoted Θ (Figure 2.1 (a)), as the position where topoIV is said to act on the DNA. By fixing this region at a site, the SAPs become rooted SAPs. At this fixed site, two segments of the polygon are 'pinched' together in order to model the gate and transported segment of the DNA being brought close together (review Figure 1.4 for reference). To model the movement of the transported segment passing through the gate segment, a strand passage is performed on the SAP by replacing the fixed Θ structure with a second structure Θ_S (Figure 2.1 (b)). This action of performing a strand passage in the SAP has the ability to change the knot type of the SAP, just as topoIV's action does.

Definition 2.12 (Θ)**.** *The fixed structure, Θ, is a specific pattern of vertices, $\mathcal{V}(\Theta)$, and edges, $\mathcal{E}(\Theta)$, cf. Figure 2.1 (a). Specifically in terms of x, y, z coordinates:*

$$\mathcal{V}(\Theta) = \{(-1,0,0), (0,0,0), (1,0,0), (0,-1,-2), (0,-1,-3), (0,0,-3), (0,1,-3), (0,1,-2)\},$$



(*a*) (*b*)

Figure 2.1: (a) The strand passage Θ structure. Closed circles represent the required vertices and closed bonds the edges. Open circles and dashed lines represent vertices and edges that Θ-SAP must not occupy for a successful strand passage to be possible. A Θ-SAP with all these vertices open is called a Θ_0-SAP. Vertex B = (0,0,0). (b) The after-strand-passage structure Θ_S. Again, the closed circles represent vertices and closed bonds represent edges. Open circles are vertices that were filled by Θ, but are now empty. Reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

and

$$\mathcal{E}(\Theta) = \{\{(-1,0,0),(0,0,0)\},\{(0,0,0),(1,0,0)\},\{(0,-1,-2),(0,-1,-3)\},$$
$$\{(0,-1,-3),(0,0,-3)\},\{(0,0,-3),(0,1,-3)\},\{(0,1,-3),(0,1,-2)\}\}.$$

The motivation for this particular Θ structure, was originally based on work by Berger *et al.* who proposed a similar shape for the topoIV-DNA structure [14].

Definition 2.13 (Θ-SAP)**.** *A SAP which contains the* Θ *structure is defined as a* Θ-SAP*, cf. Figure 2.2 (a).*

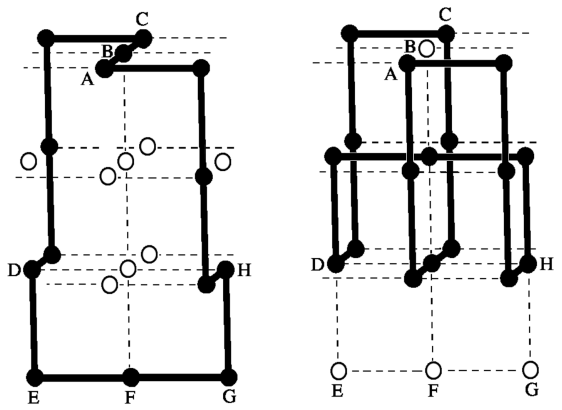
To model the strand passage topoIV performs on DNA, the Θ structure is replaced with the after-strand-passage structure Θ_S, but this is only possible when certain vertices are unoccupied. Θ-SAPs with these vertices unoccupied are called Θ_0-SAPs.

Definition 2.14 (Θ_0-SAP)**.** *A* Θ_0*-SAP is a* Θ*-SAP where all necessary vertices are free (unoccupied by the* Θ*-SAP) for a successful strand passage to occur, i.e. the open vertices seen in Figure 2.1 (a).*

Definition 2.15 (Θ_S)**.** *The* Θ_S *structure is a specific set of vertices,* $\mathcal{V}(\Theta_S)$*, and edges,* $\mathcal{E}(\Theta_S)$*, which can replace* Θ *in a* Θ_0*-SAP and result in a strand passage, cf. Figure 2.1 (b). Specifically:*

$$\mathcal{V}(\Theta_S) = \{(-1,0,0),(-1,0-1),(-1,0,-2),(0,0,-2),(1,0,-2),(1,0,-1),(1,0,0),$$
$$(0,-1,-2),(0,-1,-1),(0,0,-1),(0,1,-1),(0,1,-2)\},$$

and



(*a*) (*b*)

Figure 2.2: (a) A 14-edge Θ-SAP. This polygon would also be a Θ_0-SAP as the necessary vertices are open for a successful strand passage. (b) The corresponding after-strand-passage 18-edge polygon, Θ_S-SAP, to (a). (a) is reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

$$\mathcal{E}(\Theta_S) = \{\{(1,0,0),(1,0,-1)\},\{(1,0,-1),(1,0,-2)\},\{(1,0,-2),(0,0,-2)\},$$
$$\{(0,0,-2),(-1,0,-2)\},\{(-1,0,-2),(-1,0,-1)\},\{(-1,0,-1),(-1,0,0)\},$$
$$\{(0,-1,-2),(0,-1,-1)\},\{(0,-1,-1),(0,0,-1)\},\{(0,0,-1),(0,1,-1)\},\{(0,1,-1),(0,1,-2)\}\}.$$

Definition 2.16 (Θ_S-SAP)**.** *A SAP after a successful strand passage is called a Θ_S-SAP, as it now contains Θ_S. A Θ_S-SAP can be seen in Figure 2.2 (b).*

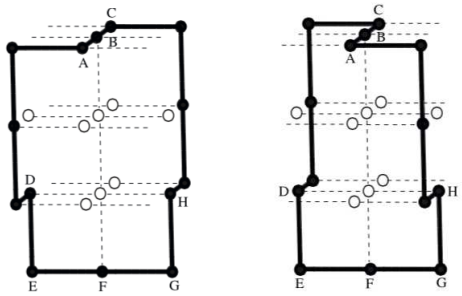
It should be noted that a Θ_0-SAP is a Θ-SAP where the Θ structure can be replaced with Θ_S, and the polygon will remain self-avoiding.

A Θ-SAP can be decomposed into the Θ structure and two SAWs which connect the two components of Θ. It is possible for the SAWs to connect to Θ in two different ways, called class I and class II. Class I means a SAW connects vertex (1,0,0) to vertex (0,-1,-2), and a SAW connects vertex (0,1,-2) to vertex (-1,0,0), or alternatively, connects A to D and H to C, such as in Figure 2.3 (a). Class II means a SAW connects vertex (-1,0,0) to vertex (0,-1,-2) and a SAW connects vertex (1,0,0) to vertex (0,1,-2), or, connects C to D and H to A, such as in Figure 2.3 (b). Szafron [105] showed that, for every Θ-SAP in class I, there exists a unique polygon in class II and vice versa. In other words there exists a mapping from one class to the other:

$$f((x,y,z)) = (-x,y,z). \tag{2.5}$$

With this mapping the two classes can be viewed as equivalent, so only one class needs to be sampled [105]. But since this mapping also reverses the sign of chiral knots, meaning a positive crossing is changed to a negative crossing (and vice versa), this must be accounted for when using a sample from one class to make conclusions about the other. By definition class II polygons have a negative crossing at Θ, while class I polygons have a positive crossing. In this work all examples will be based on class II.

With the Θ structure defined some previous polygon definitions can be refined:



(a) (b)

Figure 2.3: (a) A 14-edge Θ-SAP class I polygon. (b) A 14-edge Θ-SAP class II polygon. Reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

Definition 2.17 (p_n^{Θ}, $p_n^{\Theta_0}$, $p_n^{\Theta}(K)$, $p_n^{\Theta_0}(K)$)**.** *Define p_n^{Θ} as the number of n-edge Θ-SAPs. Define $p_n^{\Theta_0}$ as the number of n-edge Θ_0-SAPs. Define $p_n^{\Theta}(K)$ to be the number of n-edge Θ-SAPs with knot type K. Define $p_n^{\Theta_0}(K)$ to be the number of n-edge Θ_0-SAPs with knot type K.*

Here, we are interested in modelling a topoIV move and how a single strand passage can unknot a DNA knot. Therefore, we are looking at the knot type of the Θ-SAP before and after a strand passage occurs, to discover if the knot type of the Θ-SAP changes. That is, we are interested in the probability of transitioning from one knot type to another via a strand passage, which is called the *knot transition probability*.

Definition 2.18 (Knot Transition Probability)**.** *Define the* knot transition probability *from knot type K to knot type K' in an n-edge Θ-SAP to be:*

$$\rho_n^{\Theta_0}(K \to K') = \frac{p_n^{\Theta_0}(K \to K')}{p_n^{\Theta_0}(K)}. \tag{2.6}$$

where $p_n^{\Theta_0}(K)$ is the number of n-edge Θ_0-SAPs with knot type K and $p_n^{\Theta_0}(K \to K')$ is the number of these which yield a Θ_S-SAP with knot type K' after a strand passage.

Definition 2.19 (Successful Strand Passage Probability)**.** *Define the* successful strand passage probability *for an n-edge Θ-SAP with knot type K to be:*

$$\rho_n^{\Theta_0}(K \to s) = \frac{p_n^{\Theta_0}(K)}{p_n^{\Theta}(K)}. \tag{2.7}$$

Cheston *et al.* [20] used the conjectured form for the number of n-edge polygons with a fixed knot type K (Equation 2.4) to further conjecture for n-edge Θ-SAPs with knot type K as $n \to \infty$:

$$p_n^{\Theta_0}(K) \sim B_{\Theta_0 K} A_K n^{\alpha_0 + f_K + 1} \mu_0^n \tag{2.8}$$

and

$$p_n^{\Theta_0}(K \to K') \sim B_{\Theta_0(K,K')} A_{(K,K')} n^{\alpha_0 + f_K - f_{(K,K')} + 1} \mu_0^n \tag{2.9}$$

where the subscripts represent the fact that the constants are expected to depend on properties of the fixed structure and the knots listed.

Typically, a large polygon with knot type K, where K has prime knot decomposition $K = K_1 \# K_2 \# \cdots K_{f_K}$, is a large unknotted polygon with relatively localized embeddings of each of the f_K prime knots $K_1, \ldots, K_{f_K}$ randomly distributed throughout the polygon [20, 55]. So in general, there are two cases to consider for a strand passage:

Case 1: when the knot transition from $K \to K'$ alters one or more of the prime components of K

Case 2: when the knot transition from $K \to K'$ does not alter any of the prime components of K.

In this work all starting knot types K, apart from ϕ, are prime knots, so the knot decomposition consists of only one prime knot, K, so $f_K = 1$ (except when $K = \phi$, then $f_K = 0$).

These two cases result from the position of Θ relative to the 'knotted part' of the SAP. Since there is at most one original prime knot, if Θ interacts with the original prime knot then the original knot may no longer be in the prime decomposition of the resulting knot after a strand passage. If K is affected by the strand passage, this will be defined to be *Case 1*, and so $K \to K'$, $K' \neq K$. If Θ does not interact with the original prime knot, but instead is in an unknotted section of the SAP, then the original knot will be unaffected, but a new knot can be added to the resulting knot decomposition to create a new overall knot $K' = K \# K''$. This result will be defined as *Case 2* and a composite knot will be created, $K \to K' = K \# K''$ where K'' is the newly created knot. Note *Case 2* includes the possibility that K'' is an unknot, and hence includes the case $K \to K$, *i.e.* K stays the same knot type after strand passage.

For example, if $K = \phi$, then *Case 2* will always hold as there is no original prime knot that can be changed. The result can be: $K' = \phi$, if the strand passage keeps the knot type ϕ, or K' can be a newly created prime knot. Another example with knot type $K = 3_1^+$ showing both cases can be seen in Figure 2.4.

With these two cases defined, building on Equation 2.6 for the knot transition probabilities and the conjectured Equations 2.8 and 2.9, Cheston *et al.* [20] conjectured that the knot transition probabilities for these two cases become:

For *Case 1*:

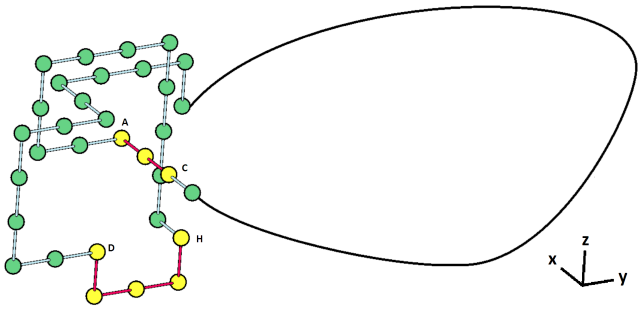
$$\rho_n^{\Theta_0}(K \to K') = \frac{p_n^{\Theta_0}(K \to K')}{p_n^{\Theta_0}(K)} \sim \frac{B_{\Theta_0(K,K')} A_\phi n^{\alpha_0 + f_K - f_{(K,K')} + 1} \mu_0^n}{B_{\Theta_0} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \left(\frac{B_{\Theta_0(K,K')} A_\phi}{B_{\Theta_0} A_K} \right) \frac{1}{n^{f_K}}, \tag{2.10}$$

where $f_{(K,K')} = 1$ as one prime component of the original knot changed, so as $n \to \infty$, $\rho_n^{\Theta_0}(K \to K') \to 0$ like n^{-f_K}.

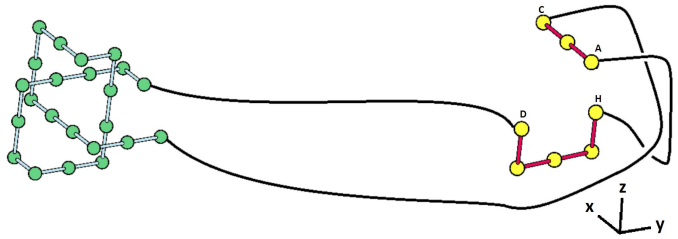
For *Case 2*:

$$\rho_n^{\Theta_0}(K \to K') = \frac{p_n^{\Theta_0}(K \to K')}{p_n^{\Theta_0}(K)} \sim \frac{B_{\Theta_0(\phi,K'')} A_K n^{\alpha_0 + f_{(\phi,K'')} + 1} \mu_0^n}{B_{\Theta_0} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \frac{B_{\Theta_0(\phi,K'')}}{B_{\Theta_0}} \equiv C_{K,K'}^{\Theta} = C_{\phi,K'}^{\Theta} \tag{2.11}$$

where $f_{(\phi,K'')} = f_K$ as no prime component of the original knot is changed, so as $n \to \infty$, $\rho_n^{\Theta_0}(K \to K') \to C_{K,K'}^{\Theta}$, a nonzero positive constant independent of the starting K.



(a)



(b)

Figure 2.4: A *Case 1* (a) and *Case 2* (b) example starting with $K = 3_1^+$. (a) *Case 1* example: Θ, the yellow vertices and red edges, can be seen to be interacting with the existing 3_1^+ knot. In this case the original 3_1^+ knot will change, resulting in a new prime knot K'. In this example the new prime knot is 0_1 so $3_1^+ \to 0_1$ (b) *Case 2* example: Θ can be seen not interacting with the existing 3_1^+ knot, and so in the unknotted portion of the polygon. In this case the original 3_1^+ will not be affected, but an additional knot can be created at Θ, resulting in a composite knot $K' = 3_1^+ \# K''$. In this particular example $K'' = 3_1^+$, so $3_1^+ \to 3_1^+ \# 3_1^+$

Additionally, for the successful strand passage probability, it is conjectured that:

$$\rho_n^{\Theta}(K \to s) = \frac{p_n^{\Theta_0}(K)}{p_n^{\Theta}(K)} \sim \frac{B_{\Theta_0} A_K n^{\alpha_0 + f_{(K)} + 1} \mu_0^n}{B_{\Theta} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \frac{B_{\Theta_0}}{B_{\Theta}} \equiv C_{K,s}^{\Theta} = C_{\phi,s}^{\Theta}, \tag{2.12}$$

so as $n \to \infty$, $\rho_n^{\Theta}(K \to s) \to C_{K,s}^{\Theta}$, a nonzero positive constant independent of the starting K.

This means that for *Case 1*, as the lengths of the polygons increase to infinity, the knot transition probability decreases to zero. What this implies is that as the length of the polygons get larger, it is less likely that Θ will interact with the existing knot. What becomes more likely, as the polygon lengths increase, is that Θ will not interact with the original existing knot, leaving the existing knot intact and possibly creating a new knot, becoming *Case 2*. For *Case 2*, as the polygon lengths increase to infinity, the knot type of the original knot does not affect the knot transition probability, which will go to a nonzero positive constant, dependent only on the newly created knot type.

We now have a model in order to model topoisomerase-DNA interactions. The CMC Θ-BFCACF algorithm (*cf.* Section 3.4) was created to study the LSP model, and hence study how a single strand passage can change the knot type of a Θ-SAP in terms of the knot transition probabilities $\rho_n^{\Theta_0}(K \to K')$ as $n \to \infty$. The conjectures in Equations 2.10-2.12 have been numerically verified with these studies [20, 108]. Additionally, this model has been used to study the local geometry of the two juxtaposed strands at the strand passage site

in order to investigate how these geometries affect the knot reduction factor and knot transition probabilities [107].

The LSP model and the associated conjectured form for the limiting value of the knot transition probability as $n \to \infty$ can be adapted to model site-specific recombination. These adaptions are presented next.

2.3 The Local Strand Exchange (LSE) Model

The LSP Model represents topoisomerase's action on DNA. In order to study recombinase's action on DNA, specifically XerCD-*dif*-FtsK, a similar model is desired. In this section this new model will be introduced, along with the associated link transition probabilities.

This new recombinase-DNA model, like the LSP model, uses the fixed Θ structure as the region where recombinase is said to act. Both recombinase and topoisomerase require two strands of the DNA to be close to one another, or 'pinched' together, a condition the fixed Θ structure satisfies. Since algorithms and analysis methods have been developed for studying Θ in the LSP model (*cf.* Chapter 3), using Θ in the recombinase model is a useful first-step. Additionally, *dif* has been found to be a preferred site for the topoIV-DNA interaction [8, 52], showing that biologically both actions can be performed at the same site. Using the same fixed structure for both models also allows for a comparison of the two models.

To develop a model for recombinase, the Θ structure now defines the two recombination sites within the polygon in close proximity, as required for the first step of the recombination process. Now, instead of a topoisomerase strand passage, we are now modelling a recombinase strand exchange, so the model is called the Local Strand Exchange (LSE) Model. Instead of replacing Θ with Θ_S, there are now two possibilities for strand exchanges: a direct-repeat-to-link strand exchange (DLE) and an inverted-repeat-to-knot strand exchange (IKE).

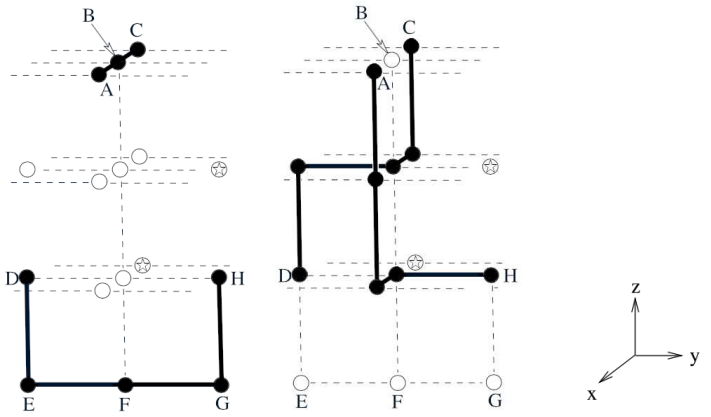
Definition 2.20 (Θ_1-SAP)**.** *Define a Θ_1-SAP to be a Θ-SAP where all necessary vertices are free (unoccupied by the Θ-SAP) for a DLE or IKE to occur, i.e. the same open vertices necessary for a successful strand passage, except for the vertex (-1,0,-2) which may now be occupied (compare Figure 2.5 (b) and Figure 2.6 (b) to see this).*

Definition 2.21 (Θ_{DLE})**.** *Θ_{DLE} is the specific set of vertices, $\mathcal{V}(\Theta_{DLE})$, and edges $\mathcal{E}(\Theta_{DLE})$, which replaces Θ in a class II Θ_1-SAP to perform a direct-repeat-to-link strand exchange, cf. Figure 2.5. Specifically:*

$$\mathcal{V}(\Theta_{DLE}) = \{(-1,0,0),(-1,0,-1),(0,0,-1),(0,-1,-1),(0,-1,-2),$$
$$(1,0,0),(1,0,-1),(1,0,-2),(0,0,-2),(0,1,-2)\},$$

and

$$\mathcal{E}(\Theta_{DLE}) = \{\{(-1,0,0),(-1,0,-1)\},\{(-1,0,-1),(0,0,-1)\},\{(0,0,-1),(0,-1,-1)\},\{(0,-1,-1),(0,-1,-2)\},$$
$$\{(1,0,0),(1,0,-1)\},\{(1,0,-1),(1,0,-2)\},\{(1,0,-2),(0,0,-2)\},\{(0,0,-2),(0,1,-2)\}\}.$$



(a) (b)

Figure 2.5: (a) Θ structure as before. Closed circles represent the required vertices and closed bonds the edges. Open circles and dashed lines represent vertices and edges that Θ-SAP must not occupy for a successful DLE. Stared vertices are vertices required for a strand passage but not for a DLE. (b) The after strand exchange structure Θ_{DLE}. Closed circles represent vertices and closed bonds represent edges. Open circles are vertices that were filled by Θ but are now empty. (a) is reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

Definition 2.22 (Θ_{IKE})**.** *Θ_{IKE} is the specific set of vertices, $\mathcal{V}(\Theta_{IKE})$, and edges $\mathcal{E}(\Theta_{IKE})$, which replaces Θ in a class II Θ_1-SAP to perform an inverted-repeat-to-knot strand exchange, cf. Figure 2.6. Specifically:*

$$\mathcal{V}(\Theta_{IKE}) = \{(-1,0,0),(-1,0,-1),(0,0,-1),(0,1,-1),(0,1,-2),$$
$$(1,0,0),(1,0,-1),(1,0,-2),(0,0,-2),(0,-1,-2)\},$$

and

$$\mathcal{E}(\Theta_{IKE}) = \{\{(-1,0,0),(-1,0,-1)\},\{(-1,0,-1),(0,0,-1)\},\{(0,0,-1),(0,1,-1)\},\{(0,1,-1),(0,1,-2)\},$$
$$\{(1,0,0),(1,0,-1)\},\{(1,0,-1),(1,0,-2)\},\{(1,0,-2),(0,0,-2)\},\{(0,0,-2),(0,-1,-2)\}\}.$$

An example of both strand exchanges performed on a class II 14-edge Θ-SAP can be seen in Figure 2.7.

With DLE and IKE defined, some earlier definitions need to be expanded to include the new model. In order to differentiate between the strand passage and the general strand exchange, SP and SE superscripts will be used. When considering a specific strand exchange, DLE or IKE superscripts will be used as appropriate. Also, recall that knots are a subset of links so 'links' also include knots.

Definition 2.23 ($p_n^{\Theta_1}, p_n^{\Theta_1}(K)$)**.** *Define $p_n^{\Theta_1}$ to be the number of n-edge class II Θ_1-SAPs and define $p_n^{\Theta_1}(K)$ to be the number of these with knot type K.*

Definition 2.24 (Link Transition Probability)**.** *Define the* link transition probability *from a knot type K to link type L in an n-edge class II Θ-SAP to be:*

$$\rho_n^{SE}(K \to L) = \frac{p_n^{SE}(K \to L)}{p_n^{\Theta_1}(K)}, \tag{2.13}$$

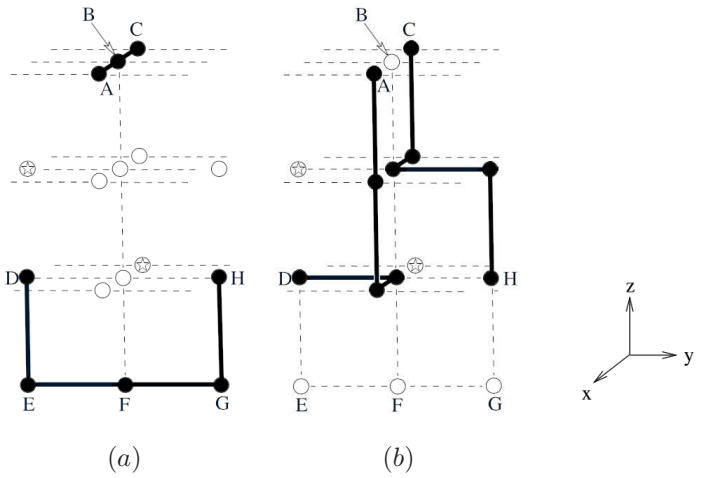


Figure 2.6: (a) Θ structure as before. Again closed circles represent the required vertices and closed bonds the edges. Open circles and dashed lines represent vertices and edges that Θ-SAP must not occupy for a successful IKE. Stared vertices are vertices required for a strand passage but not for an IKE. (b) The after strand exchange structure Θ_{IKE}. Closed circles represent vertices and closed bonds represent edges. Open circles are vertices that were filled by Θ but are now empty. (a) is reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

where as before, $p_n^{\Theta_1}(K)$ *is the number of n-edge class II* Θ_1*-SAPs with knot type K and* $p_n^{SE}(K \to L)$ *is the number of these which yield link type L.*

Definition 2.25 (Successful Strand Exchange Probability)**.** *Define the* successful strand exchange probability *for an n-edge class II* Θ*-SAP with knot type K to be:*

$$\rho_n^{SE}(K \to s) = \frac{p_n^{\Theta_1}(K)}{p_n^{\Theta_{II}}(K)}, \tag{2.14}$$

where $p_n^{\Theta_{II}}(K)$ is the number of n-edge class II Θ-SAPs with knot type K.

The reason for two separate strand exchanges is to account for both direct and inverted repeats of the *dif* sites in the DNA, as indicated by the strand exchange names. Recall, when the *dif* sites are in direct repeat, the result is a link with two components (*cf.* Figure 1.9 (a)). Figure 2.8 (a) shows a cartoon version of a polygon with direct repeats, which contains the Θ structure in red, and two SAWs in black connecting Θ to give a class II Θ_0-SAP. In this picture the blue arrows show the recombination sites are in direct repeat. When a direct-repeat-to-link strand exchange (DLE) occurs (Figure 2.8 (b)) the red Θ is replaced with the purple Θ_{DLE}, resulting in the creation of two separate polygons and a link. The resulting link type will depend on the specifics of the two SAWs. It can be seen that a DLE is equivalent to changing the crossing at Θ from a negative crossing (for class II polygons) to a positive crossing (when projected down the x-axis).

Recall, when the *dif* sites are in inverted repeat the result is a single component knot (*cf.* Figure 1.9 (b)). Figure 2.9 (a) shows a cartoon version of a polygon with inverted repeats. When an IKE strand exchange occurs, (Figure 2.9 (b)) the red Θ is replaced with the purple Θ_{IKE}, resulting in a new knot. Again, the knot type depends on the two SAWs, and an IKE move is equivalent to changing the negative crossing at Θ (for class II polygons) to no crossing (when projected down the x-axis).

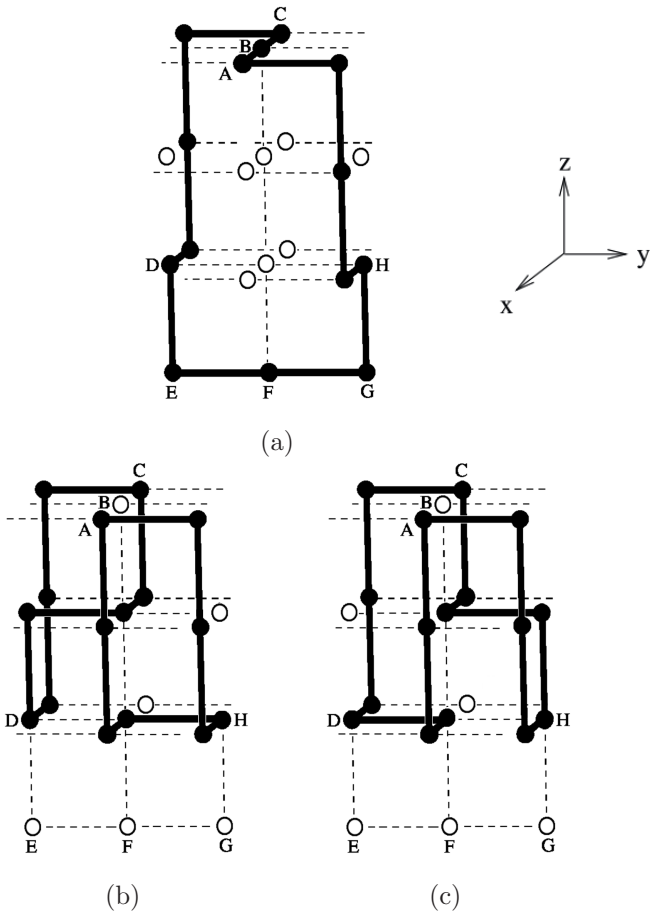


Figure 2.7: (a) A 14-edge class II Θ-SAP. (b) The corresponding after direct-repeat-to-link strand exchange polygon. DLE on a class II polygon results in a link, or two separate polygons which may or may not be linked together. (c) The corresponding after inverted-repeat-to-knot strand exchange polygon. An IKE exchange on a class II polygon results in a single knotted polygon. (a) is reproduced from [20] in accordance with the 'Author Rights' section 3.2.2 of the IOP copyright agreement.

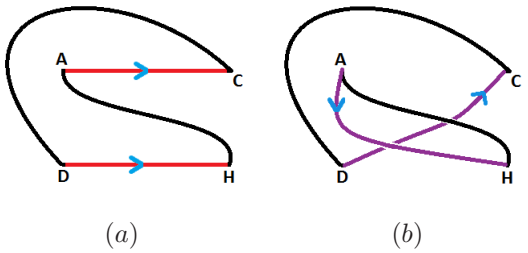
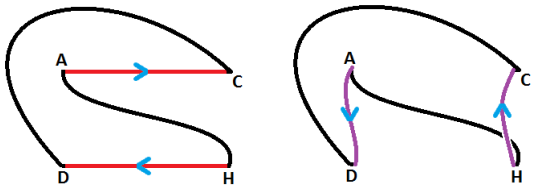


Figure 2.8: (a) A cartoon version of a polygon with direct repeats. The red represents the Θ structure, the black is two SAWs, and the blue arrows the direction of the recombination sites. (b) The corresponding polygon after a DLE, where the red Θ is replaced with the purple Θ_{DLE}, creating two separate polygons, and so a link.



(a) (b)

Figure 2.9: (a) A cartoon version of a polygon with inverted repeats. The red represents the Θ structure, the black is two SAWs, and the blue arrows the direction of the recombination sites. (b) The corresponding polygon after an IKE, where the red Θ is replaced with the purple Θ_{IKE}, creating a new knot.

The above definitions are restricted to class II Θ-SAPs. However, for a given knot type K, to get a complete picture of the effect of a strand exchange operation on the full set of Θ-SAPs, class I Θ-SAPs should also be considered. Equation 2.5 defines a *mirror* operation that allows the use of class II Θ-SAPs to also study class I Θ-SAPs. For each class II Θ-SAP ω, there is a unique class I Θ-SAP given by $mirror(\omega)$. Thus, the probability that an n-edge class I knot type K Θ-SAP will have link type L after a $mirror$(SE) operation, denoted $\rho_n^{mirror(SE)}(K \to L)$, is equal to $\rho_n^{SE}(mirror(K) \to mirror(L))$. In the case that K is an achiral knot, $mirror(K) = K$, and hence the class I results are obtained directly from those for class II by determining $mirror(L)$. For chiral K, the class I results are obtained by performing SE on Θ-SAPs with knot type $mirror(K)$, the mirror image of knot type K. For example, to obtain class I results for $K = 3_1^+$, class II knot type 3_1^- SAPs are studied. As argued in Cheston *et al.* [20] for SP, for the case of chiral knots K, in order to combine the results for the two classes of polygons to obtain a single link transition probability, one would need to know, for example, $\frac{p_n^{\Theta_{II}}(K)}{p_n^{\Theta_{II}}(mirror(K))}$. It is expected that asymptotically $p_n^{\Theta_{II}}(K) \sim p_n^{\Theta_{II}}(mirror(K))$, in which case the two classes of polygons can be weighted equally to determine link transition probabilities; however, for small lengths n, equality is not expected. In this thesis, the ratio $\frac{p_n^{\Theta_{II}}(K)}{p_n^{\Theta_{II}}(mirror(K))}$ is not studied, hence, when results for the two classes are combined it is assumed that they are equally likely, as is expected to be the case for large n.

For the case of DLE, using tangle analysis there are three tangle scenarios (as shown in Figure 2.10 which details the last three unlinking steps of Figure 1.28) that all correspond to the same recombinase action resulting in a link: one scenario introduces a positive crossing, denoted (+1), a second introduces a negative crossing, denoted (-1), and the third introduces no crossing, denoted (0). Vazquez *et al.* [115] proved that these three senarios in Figure 2.10 can all be obtained from the same 3-dimensional DNA molecule (before and after a recombinase action) by taking different 2-dimensional projections. Interestingly, because the Θ structure is fixed in space, when the focus is on a fixed projection along the $x = y$ axis, DLE on a class II Θ-SAP fits with the (+1) scenario, and $mirror$(DLE) on a class I Θ-SAP fits with the (0) scenario. Even though projections in other directions would yield the other scenarios, the use of this $x = y$ projection gives a convenient way to refer to the two classes of Θ-SAPs. That is, the (+1) will refer to class II Θ-SAPs and

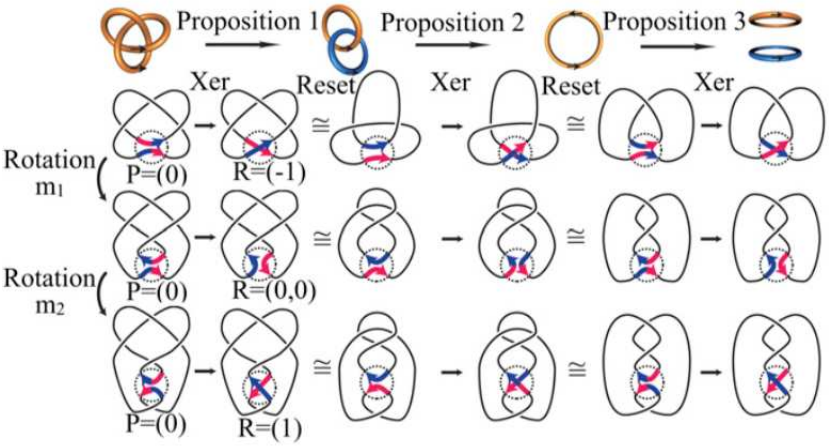


Figure 2.10: A detailed picture of the unlinking path from 3_1^+ to the unlink. It shows there are three possible scenarios of a Xer move, R=(+1), R=(-1) and R=0, depending on what the original knot projection is. Reproduced from [95] in accordance with http://www.pnas.org/site/aboutpnas/rightperm.xhtml.

DLE, while the (0) scenario will refer to class I Θ-SAPs and *mirror*(DLE).

For chiral knots:

$$\rho_n^{(+1)}(K^+ \to L) = \rho_n^{DLE}(K^+ \to L) \tag{2.15}$$

$$\rho_n^{(0)}(K^+ \to L) = \rho_n^{mirror(DLE)}(K^+ \to L) = \rho_n^{DLE}(K^- \to mirror(L)) \tag{2.16}$$

The same equations apply for K^- knots.

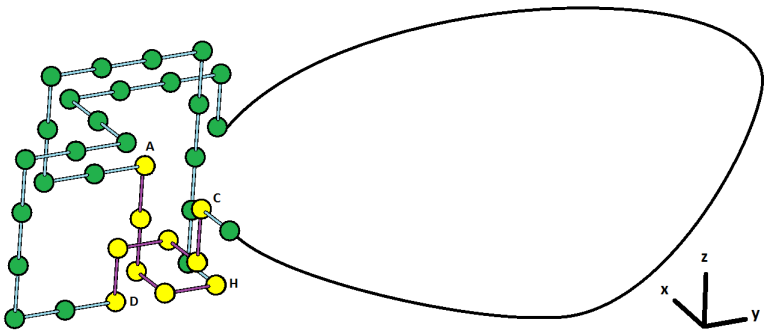
For achiral knots:

$$\rho_n^{(+1)}(K \to L) = \rho_n^{DLE}(K \to L) = \rho_n^{mirror(DLE)}(K \to mirror(L)) = \rho_n^{(0)}(K \to mirror(L)) \tag{2.17}$$

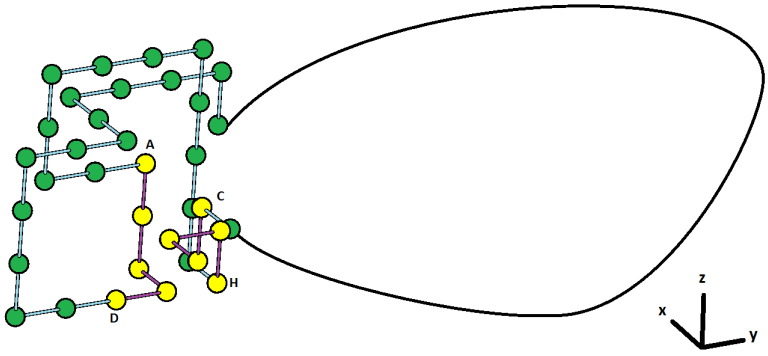
Similarly for IKE, $p_n^{mirror(IKE)}(K \to L) = p_n^{IKE}(mirror(K) \to mirror(L))$.

When it is assumed that both polygon classes are equally likely for knot type K, this property is used in this thesis in order to exclusively look at class II polygons but gain all the recombinase information.

Like with the LSP Model, for the LSE model there are two cases for the position of Θ relative to the 'knotted part' of the Θ-SAP: interacting with the existing knot or not interacting with the existing knot. In *Case 1*, Θ interacts with the original knot such that the original knot is no longer in the prime knot



(a)

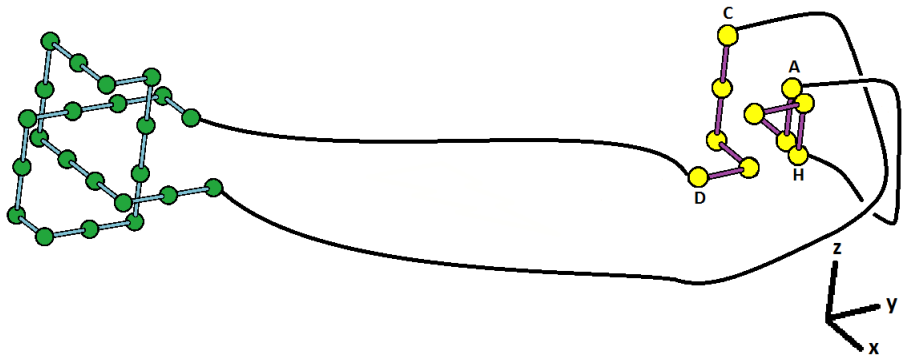


(b)

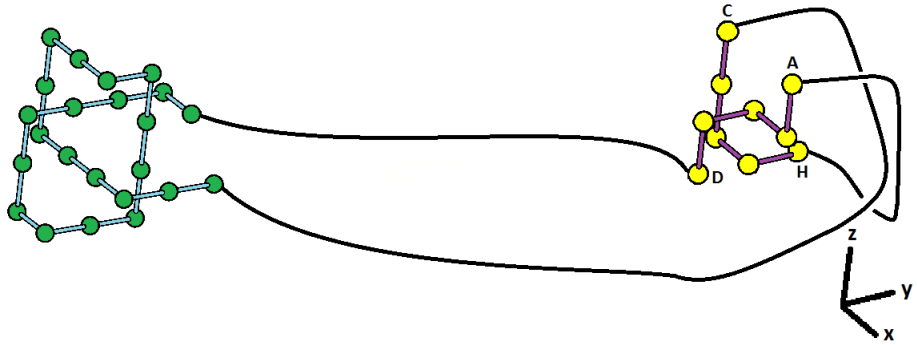
Figure 2.11: *Case 1* SE example. The Θ structure, the yellow vertices and red edges, can be seen to interact with the existing 3_1^+ knot, so the original knot will change. (a) A DLE will create a link at Θ resulting in $K \to L$ where $L = K' \cup K''$. In this example the result is $3_1^+ \to 2_1^2$, also known as the Hopf link. (b) An IKE will create knot at Θ, resulting in a new prime know $K \to K'$. In this example the new prime knot is ϕ so $3_1^+ \to \phi$.

decomposition of the link after a strand exchange. The outcome of a DLE in this case is the creation of a two component link, L, either linked or unlinked, $K \to L$, where $L = K' \cup K'', K' \neq K, K'' \neq K$. The outcome of an IKE in this case is similar to the strand passage outcome, the original knot is not in the prime knot decomposition of the resulting knot after an IKE at Θ, $K \to K', K' \neq K$. Note that when $K = \phi$ *Case 1* is not possible since there is no original knot that can no longer be in the decomposition of the link after a strand exchange. An example of *Case 1* with knot type $K = 3_1^+$ can be seen in Figure 2.11.

If Θ does not interact with the original knot, and so in *Case 2*, the original knot will remain intact. The outcome of a DLE is the creation of a two component link at Θ, where the knot decomposition of one component of the link contains the original knot K. The resulting two components of the link may or may not be linked together. If the two knots are linked together they will be represented as $L = l(K' \cup K'')$, or if unlinked as $L = u(K' \cup K'')$, where K is within the knot decomposition of one of the components. The outcome of an IKE is similar to the strand passage outcome, a composite knot will be created with the original knot in the prime decomposition of the resulting knot, $K \to K'$ where $K' = K \# K''$. An example of *Case 2* with knot type $K = 3_1^+$ can be seen in Figure 2.12.



(a)



(b)

Figure 2.12: *Case 2* example. The Θ structure, the yellow vertices and red edges, can be seen not interacting with the existing 3_1^+ knot, meaning the original knot will remain intact. (a) A DLE will create a new knot at Θ resulting in an overall link, $K \to L$ where $L = K \cup K'$. In this example the result is $3_1^+ \to l(3_1^+ \cup \phi)$. (b) An IKE will create a new knot at Θ resulting in an overall composite knot $K \to K'$ where $K' = K \# K''$. In this example $K'' = \phi$, so $3_1^+ \to 3_1^+ \# \phi = 3_1^+$.

The conjectured forms for the knot transition probabilities for the two cases in the LSP model (Equations 2.10 and Equation 2.11) also apply for the LSE model with a DLE and an IKE taken into account. Therefore, they become:

For *Case 1* DLE:

$$\rho_n^{DLE}(K \to L) = \frac{p_n^{DLE}(K \to L)}{p_n^{\Theta_1}(K)} \sim \frac{B_{DLE(K,L)} A_L n^{\alpha_0 + f_K - f_{(K,L)} + 1} \mu_0^n}{B_{\Theta_1} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \left(\frac{B_{DLE(K,L)} A_\phi}{B_{\Theta_1} A_L} \right) \frac{1}{n^{f_K}}, \quad (2.18)$$

where $f_{(K,L)} = 1$ as one prime component of the original knot is changed, so as $n \to \infty$, $\rho_n^{DLE}(K \to L) \to 0$ like n^{-f_K}.

For *Case 1* IKE (similar to Equation 2.10):

$$\rho_n^{IKE}(K \to K') = \frac{p_n^{IKE}(K \to K')}{p_n^{\Theta_1}(K)} \sim \frac{B_{IKE(K,K')} A_\phi n^{\alpha_0 + f_K - f_{(K,K')} + 1} \mu_0^n}{B_{\Theta_1} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \left(\frac{B_{IKE(K,K')} A_\phi}{B_{\Theta_1} A_K} \right) \frac{1}{n^{f_K}}, \quad (2.19)$$

where $f_{(K,K')} = 1$ as one prime component of the original knot is changed, so as $n \to \infty$, $\rho_n^{IKE}(K \to K') \to 0$, like n^{-f_K}.

For *Case 2* DLE:

$$\rho_n^{DLE}(K \to L) = \frac{p_n^{DLE}(K \to L)}{p_n^{\Theta_1}(K)} \sim \frac{B_{DLE(\phi,L)} A_L n^{\alpha_0 + f_{(\phi,L)} + 1} \mu_0^n}{B_{\Theta_1} A_L n^{\alpha_0 + f_K + 1} \mu_0^n} = \frac{B_{DLE(\phi,L)}}{B_{\Theta_1}} \equiv C_{K,L}^{DLE} = C_{\phi,L}^{DLE}, \quad (2.20)$$

where $f_{(\phi,L)} = f_K$ as no prime component of the original knot is changed, so as $n \to \infty$, $\rho_n^{DLE}(K \to L) \to C_{K,L}^{DLE}$, a constant independent of the starting K.

For *Case 2* IKE (similar to Equation 2.11):

$$\rho_n^{IKE}(K \to K') = \frac{p_n^{IKE}(K \to K')}{p_n^{\Theta_1}(K)} \sim \frac{B_{IKE(\phi,K'')} A_K n^{\alpha_0 + f_{(\phi,K'')} + 1} \mu_0^n}{B_{\Theta_1} A_K n^{\alpha_0 + f_K + 1} \mu_0^n} = \frac{B_{IKE(\phi,K'')}}{B_{\Theta_1}} \equiv C_{K,K'}^{IKE} = C_{\phi,L}^{IKE}, \quad (2.21)$$

where $f_{(\phi,K'')} = f_K$ as no prime component of the original knot is changed, so as $n \to \infty$, $\rho_n^{IKE}(K \to K') \to C_{K,K''}^{IKE}$, a constant independent of the starting K.

Comparing Equations 2.10 and 2.11 to Equations 2.18-2.21, it can be seen that the LSP and LSE have the same pattern: as the polygon length increases to infinity it is conjectured to become more likely that Θ will not interact with the existing knot, so either an additional knot or a link will be created at Θ leaving the original knot intact, *i.e. Case 1* link transition probabilities both go to zero, while *Case 2* link transition probabilities go to nonzero positive constants, independent of the starting knot type K. In Chapter 4 these conjectures will be explored numerically.

2.4 Chapter 2 Summary

In this chapter self-avoiding polygons (SAPs) were introduced as the method used to model DNA in this thesis. It was reviewed that as the length of a polygon goes to infinity, the probability that the polygon will be knotted goes to one, while the probability that it will have a fixed knot type K goes to zero. A conjectured asymptotic form for the number of fixed knot type polygons was then presented in Equation 2.4. Next, the Local Strand Passage (LSP) model was reviewed as the method used to model topoisomerase action with the use of the fixed Θ structure. It was shown that there are two cases for the location of the knotted part within the SAP relative to Θ, each giving different knot transition probabilities. For *Case 1*, Θ interacts with the original knot of the polygon so a strand passage changes this knot, and for *Case 2* Θ does not interact with the original knot so a strand passage can create a new knot at Θ, forming an overall composite knot. Conjectured forms for these probabilities were given in Equations 2.10 and 2.11, showing that *Case 1*'s probability goes to zero and *Case 2*'s probability goes to a constant independent of the starting knot K. Finally, the new Local Strand Exchange (LSE) model was introduced as the method used to model recombinase action, also through the use of Θ. It was shown that two strand exchanges are possible, the direct-repeat-to-link strand exchange (DLE) and the inverted-repeat-to-knot strand exchange (IKE). Like the LSP model, it was shown that there are two cases for the location of the knotted part of the SAP, interacting or not interacting with the Θ. Conjectured forms were then presented in Equations 2.18-2.21. Again, it is conjectured that *Case 1*'s probability goes to zero as the length of the SAPs go to infinity, while *Case 2*'s probability goes to a nonzero positive constant independent of the starting knot K, for each strand exchange. In Chapter 4, these conjectured forms will be explored numerically using a similar approach to that used for the LSP model in [89, 105, 106].

Chapter 3

Algorithm and Analysis Techniques

This chapter is devoted to the algorithms used to generate Θ-SAPs, and the analysis techniques used in order to study link transition probabilities. Once the algorithms used are explained, some specifics of the simulation will be given, followed by a review of a number of analysis techniques used on the simulation data. This chapter relies heavily on theory, algorithms and techniques developed previously by others for the study of Θ-SAPs. Because of this, only an overview will be provided on how they apply to the Local Strand Exchange (LSE) model. For full details, the reader will be referred to the original work.

The algorithm used to generate the Θ-SAPs studied here is called the Θ-BFACF algorithm and was developed by Szafron and Soteros [105, 106, 107, 108]. The Θ-BFACF algorithm is a Markov chain Monte Carlo algorithm that is based on the original BFACF algorithm [13, 18, 19]. Before an explanation of this algorithm can be given, some Markov chain and Monte Carlo terminology and theory must first be defined and explained.

3.1 Markov Chains and Monte Carlo Methods: Definitions and Theory

Definition 3.1 (Stochastic Process [85]). *A* stochastic process $\{X_t, t \in T\}$ *is a collection of random variables. That is, for each* $t \in T$, X_t *is a random variable.* T *is called the index set of the process, and can be discrete if* T *is countable, or continuous if* T *is an interval on the real line.*

In this thesis, T is discrete and indexed by non-negative integers and we will be considering the elements of T as *time steps* of a simulation.

Definition 3.2 (State, State Space [85]). *The random variable* X_t *is called the* state *at time* t *of the stochastic process* $\{X_t, t \in T\}$. *The set of all values that the random variables* X_t *can assume is the* state space, $\mathcal{S}$.

Definition 3.3 (Markov Process [56]). *A* Markov process *is a process,* $\{X_t, t \in T\}$, *with state space* $\mathcal{S}$, *that has the property: given the value for* X_{t_n}, *the values for* X_t, *for all* $t > t_n$, *do not depend on the values of* X_u, *for all* $u < t_n$. *That is, the future behaviour is not dependent on past states; it is only dependent on the*

current state. Formally:

$$P\{a < X_t \leq b | X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_n} = x_n\} = P\{a < X_t \leq b | X_{t_n} = x_n\},$$
$$\textit{whenever } t_1 < t_2 < \cdots < t_n < t, \quad x_i \in \mathcal{S} \text{ for } i = 1, \ldots, n.$$

A Markov process that has a finite, countable state space is called a *Markov chain*. A discrete time Markov chain $\{X_t, t \in T\}$ is a Markov chain with index parameter set $T = \{0, 1, 2, \ldots\}$. For a discrete time Markov chain, Definition 3.3 reduces to stating that the conditional distribution of any future state, X_{t+1}, depends only on the present state X_t, not on any other past states. The simplest situation is when this distribution is also independent of t (time-homogeneous). In this case, let $\{X_t, t \in T\}$ be a discrete time Markov chain with state space $\mathcal{S}$. Let P_{xy} represent the one-step probability of the process transitioning from state x to state y in the next step. The n-step probability is then [85]:

$$P_{xy}^{(n)} = P\{X_{n+k} = y | X_k = x\}, \qquad n > 0, \quad x, y \in \mathcal{S}. \tag{3.1}$$

Definition 3.4 (Accessible, Communicate, Communication Class [85])**.** *State y is* accessible *from state x if $P_{xy}^{(n)} > 0$ for some $n \geq 0$. That is, state y is accessible from state x if and only if, starting in x, it is possible for the process to enter state y in a finite number of steps n. Two states which are accessible to each other are said to* communicate*, and are said to be in the same* communication class *(or* class *for short).*

Definition 3.5 (Irreducible [56, 106])**.** *A Markov chain is said to be* irreducible *if, for all $x, y \in \mathcal{S}$, x and y communicate, i.e. for each pair of $x, y \in \mathcal{S}$, $P_{xy}^{(n)} > 0$ for some $n \geq 0$.*

Definition 3.6 (Stationary Distribution [56])**.** *A set of probabilities $\pi := \{\pi_x\}_{x \in \mathcal{S}}$ is a stationary distribution of $\{X_t, t \in T\}$, if for all $x \in \mathcal{S}$,*

$$\pi_x \geq 0, \quad \sum_{x \in \mathcal{S}} \pi_x = 1, \quad \textit{and} \quad \sum_{x \in \mathcal{S}} \pi_x P_{xy} = \pi_y, \quad \textit{for all } y \in \mathcal{S}.$$

Definition 3.7 (Period, $d(x)$, Aperiodic [56, 106])**.** *The* period *of a state x, $d(x)$, is the greatest common divisor of all integers $n \geq 0$ for which $P_{xx}^{(n)} > 0$. A Markov chain is said to be* aperiodic *if for all $x \in \mathcal{S}$, $d(x) = 1$.*

Note that period is a class property.

Definition 3.8 (Recurrent [56, 106])**.** *Let $f_{xy}^{(i)}$ be the probability that it will take i transitions to first reach state y from state x. Then state x is defined to be* recurrent *if and only if*

$$\sum_{i=1}^{\infty} f_{xx}^{(i)} = 1.$$

If states x and y communicate, and x is recurrent, then y is also recurrent (*i.e.* recurrence is a class property). If all states $x \in \mathcal{S}$ are recurrent then the Markov chain is said to be recurrent. A Markov chain is *positive recurrent* if it is recurrent and there exists some $x \in \mathcal{S}$ where $\lim_{t \to \infty} P_{xx}^{(n)} > 0$. This means that the probability of returning to state x is 1 and the expected first return time is finite.

Definition 3.9 (Ergodic [56, 106]). *If a Markov chain is irreducible, positive recurrent, and aperiodic, then the Markov chain is* ergodic *in* $\mathcal{S}$.

The Markov chains studied in this thesis are ergodic [105, 106]. The following theorem is thus an important result for this work.

Theorem 4 (Equilibrium Distribution, π [56, 106]). *If* $\{X_t, t \in T\}$ *is an ergodic Markov chain, then as* $n \to \infty$, $\{X_t, t \in T\}$ *will converge to a unique* equilibrium distribution, π, *independent of the initial starting state. That is there exists a set of probabilities,* $\pi := \{\pi_x\}_{x \in \mathcal{S}}$, *such that* $\sum_{x \in \mathcal{S}} \pi_x = 1$ *and* $\lim_{t \to \infty} P(X_t = x) = \pi_x$ *for each* $x \in \mathcal{S}$. π *is also the unique stationary distribution of the Markov chain as in Definition 3.6.*

When a quantity (or quantities) is (are) computationally difficult to evaluate, a Monte Carlo method can be used to estimate the quantity (or quantities).

Definition 3.10 (Monte Carlo Method [85, 106]). *A* Monte Carlo method *is a numerical method for generating states of a random variable according to a specific probability distribution.*

Definition 3.11 (Markov chain Monte Carlo (MCMC) method [85, 106]). *A* Markov chain Monte Carlo (MCMC) method *is a Monte Carlo method using Markov chains. Instead of a Monte Carlo method generating independent random vectors, it generates a sequence of successive states of a Markov chain.*

MCMC methods are typically used to study the equilibrium distribution of an ergodic Markov chain. For an ergodic Markov chain however, unless you start from the equilibrium distribution, data generated from MCMC is not immediately in equilibrium, but is expected to reach equilibrium as the number of time steps goes to infinity. Furthermore, once in equilibrium, the data generated is not independent, so it is important to take that into account when obtaining estimates and their confidence intervals. In practice, the time to equilibrium and the time between independent data points can be estimated using autocorrelation functions. For example, Sokal [97] showed that the convergence of a chain to its equilibrium distribution is bounded above by the exponential autocorrelation time, τ_{exp}, as defined in [89, Equation 3.16]. Additionally, for a real-valued function A defined on the Markov chain, $\{X_t, t \in T\}$, we say two observed values $A(X_{t_1})$ and $A(X_{t_2})$, $t_2 > t_1 > 0$, are *essentially independent* if $|t_2 - t_1| \geq 2\tau_{int,A}$ where $\tau_{int,A}$ is the *integrated autocorrelation time* for an observable A, as defined in [89, Equation 3.27]. These two autocorrelation times need to be estimated in order to properly analyze data from a MCMC simulation. For full details and more discussion see [89, 106].

The specific MCMC method used to generate the Θ-SAPs for this work is the Θ-BFACF algorithm developed by Szafron [106]. This algorithm was based on the BFACF algorithm [13, 18, 19] which will be reviewed next. Once the BFACF and Θ-BFACF algorithms have been introduced, the main algorithm for data generation (CMC Θ-BFACF) will be introduced. Following this, techniques will be presented in order to properly analyze data from this MCMC simulation.

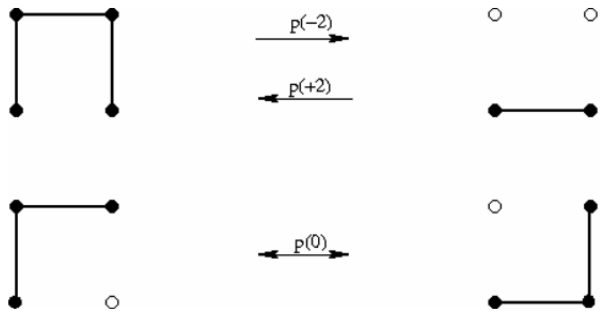


Figure 3.1: Types of BFACF moves. The top shows the p(-2) move, left to right, subtracting two edges and the p(+2) move, right to left, adding two edges. The bottom shows the p(0) which doesn't change the edge count, but flips two edges into a mirrored position. Reproduced from [106] with permission from the author.

3.2 The BFACF Algorithm

The BFACF algorithm was developed by Berg, Foerster, Aragão de Carvalho, Caracciolo and Fröhlich [13, 18, 19] to generate a Markov chain on the set of all possible self-avoiding walks with variable-lengths and fixed endpoints in $\mathbb{Z}^3$ [65]. The BFACF algorithm has been applied to SAPs [113], and it has been proven that the BFACF algorithm is ergodic on the set of all SAPs with knot type K, meaning that starting with a SAP ω with knot type K, using the BFACF algorithm, it is possible to reach any other SAP ω' with knot type K, but the BFACF algorithm cannot change the knot type of a SAP [65, 113].

The BFACF algorithm makes local changes to a starting SAP, ω, with length n, by picking a random edge of the polygon and attempting to move the edge in a direction perpendicular to itself, and adding or subtracting necessary edges to ensure ω stays connected. There are three possible moves, called BFACF moves, which can be attempted (*cf.* Figure 3.1): the p(+2) move adds two edges to the polygon; the p(-2) move subtracts two edges from the polygon; and the p(0) move keeps the edge count the same but flips two edges. After attempting the desired move, if the resulting graph, called ω', is still a self-avoiding polygon, the move is considered successful and ω becomes ω'; otherwise the state remains ω. A single time step consists of an attempted BFACF move.

Once an edge is randomly chosen, the probability that a particular BFACF move is attempted is defined by:

$$Pr(\omega') \coloneqq \begin{cases} p_n(+2) \coloneqq \frac{(n+2)^{q-1}z^2}{n^{q-1}+3(n+2)^{q-1}z^2}, & if |\omega'| - |\omega| = 2, \\ p_n(-2) \coloneqq \frac{(n-2)^{q-1}z^2}{(n-2)^{q-1}+3^{q-1}z^2}, & if |\omega'| - |\omega| = -2, \\ p_n(0) \coloneqq \frac{p_n(+2)+p_n(-2)}{2}, & if |\omega'| - |\omega| = 0, \\ 0, & otherwise \end{cases} \tag{3.2}$$

where q is an integer > 0, and z is called the fugacity of the chain.

The equilibrium distribution of the BFACF algorithm is given by [106, Equation 3.9]:

$$\pi_\omega(q,z) = \frac{|\omega|^q z^{|\omega|}}{\sum_{i=0}^{\infty} i^q p_i(K) z^i} \tag{3.3}$$

for each SAP ω with knot type K, where z is the fugacity of the chain $0 < z < z_c(K)$ and q is an integer > 0, bounded by $z_c(K)^2 \leq (\frac{2}{3})^{q-1}$. The fixed value of $z_c(K)$ is believed to be independent of K, with $z_c(K) = \frac{1}{\mu_\phi}$ where μ_ϕ is the growth constant for unknotted SAPs (Equation 2.11). Increasing z and q increases the average polygon length generated.

For a detailed discussion of the BFACF algorithm see [106] and references therein.

3.3 The Θ-BFACF Algorithm

The Θ-BFACF algorithm is a specialized BFACF algorithm developed by Szafron and Soteros [105, 106, 107, 108] to generate Θ-SAPs [105]. The Θ-BFACF algorithm is similar to the BFACF algorithm; however, it ensures that the Θ structure is not altered throughout the simulation. Therefore, if the chosen BFACF move alters Θ the move is rejected and $\omega' = \omega$.

Given a starting Θ-SAP with knot type K, the Θ-BFACF algorithm generates variable-length Θ-SAPs with knot type K in the same connection class as started. The target equilibrium distribution for a Markov chain generated by the Θ-BFACF algorithm is given by [108, Equation 21]:

$$\pi_\omega(q,z) = \frac{(|\omega|-6)|\omega|^{(q-1)} z^{|\omega|}}{\sum_{i=0}^{\infty}(i-6) i^{q-1} p_i^{\Theta}(K) z^i} \tag{3.4}$$

for each fixed class (I or II) Θ-SAP ω with knot type K, where $0 < z < z_c(K)$ and q is an integer > 0, bounded by $z_c(K)^2 \leq (\frac{2}{3})^{q-1}$[108]. For full details on the Θ-BFACF algorithm see [105, 106, 107, 108].

We are interested in modelling DNA, which is very long, so large polygons are needed. To obtain these large polygons, z must increase towards $z_c(K)$. However, a limitation of the BFACF algorithm, and hence the Θ-BFACF algorithm, is that as z approaches $z_c(K)$, the time it takes for the algorithm to reach equilibrium approaches infinity, exponentially [96]. As well, as z increases, the integrated autocorrelation time also increases, and hence the time between essentially independent data points increases. This poses a problem for obtaining statistically reliable data for the polygon sizes of interest. One method to decrease the autocorrelation times for the BFACF, and so the Θ-BFACF algorithm, is the Composite Markov Chain (CMC) Θ-BFACF method, discussed next.

3.4 Composite Markov Chain (CMC) Θ-BFACF

Composite Markov chain Monte Carlo (CMCMC) methods have been used to study SAPs [74, 75] and were adapted by Szafron [105, 106] to become the Composite Markov chain (CMC) Θ-BFACF algorithm for studying Θ-SAPs. The CMC Θ-BFACF involves the implementation of M Θ-BFACF algorithms (called chains) run in parallel, each with different z values, and state swaps are attempted between the different

chains in the system. By swapping, dependence between chains is introduced, meaning that each chain is no longer Markov; however, the whole process, called a *composite Markov chain*, has the Markov property. It's been proven that the equilibrium distribution for the CMC Θ-BFACF algorithm is $\{\prod_{i=1}^{M} \pi_{\omega_i}(q, z_i)$: for each fixed class (I or II) Θ-SAP, ω_i with knot type $K\}$, where $\pi_{\omega}(q, z_i)$ is from Equation 3.4 [106].

Chain i, $1 \leq i \leq M$, convergences to its own marginal equilibrium distribution, $\pi_{\omega_i}(q, z_i)$, usually with some overlap between the distributions of two adjacent chains. If one chain is slow to converge and one is fast, swapping between chains decreases the time it takes the whole process to reach its equilibrium distribution [106]. Swapping also prevents chains from getting stuck in a local state for a long period of time, which can lead to misleading results [89]. A single time step of this method consists of either one Θ-BFACF move attempted on each of the M chains of the CMCMC (called a Θ-BFACF move in parallel), or a single attempted swap between two chains. For full details on the CMC Θ-BFACF algorithm method see [106].

The next section will give specifics of the CMC Θ-BFACF algorithm simulations used to generate the data used in this thesis.

3.5 Implementation of CMC Θ-BFACF to Study Link Transition Probabilities

With the algorithm used to generate a sample of Θ-SAPs defined, an outline of a typical simulation will be provided, followed by the general steps taken in order to properly analyze the generated data. All the presented data analysis techniques have previously been developed by Szafron [104, 106] based on the appropriate Markov chain Monte Carlo analysis theory by Fishman [37, Chapter 6], Gelman [43], Gelman and Rubin [44] and Sokal [97]. How these techniques were applied to the LSE model will be outlined; however full details are outside the scope of this thesis and can be found in [104, 106].

All the data analyzed in this thesis was previously generated from a CMC Θ-BFACF algorithm simulation consisting of ten independent replications for each starting knot type $K = \phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+$, and 5_2^-. Each replication for each knot type was run for a total of 4.80×10^{11} time steps (4.00×10^{11} Θ-BFACF moves in parallel and 0.8×10^{11} attempted swaps), where a swap was attempted every five Θ-BFACF moves in parallel and the Θ-SAPs were sampled from each chain after every 120,000 time steps. There were 14 chains run in parallel, with chain 1 having the smallest convergence time, with $z = z_1 = 0.2030$ and chain 14 having the largest, with $z = z_{14} = 0.2132$. Note that $q = 1$ for all $K \neq \phi$ and $q = 2$ for $K = \phi$. The full set of values $z_1, z_2, \ldots, z_{14}$ is given in [107]. For $1 \leq i \leq 14$, the marginal equilibrium distribution corresponding to chain i is given by Equation 3.4 with $z = z_i$.

For each knot type, each simulation generated a sample of $560,000,000$ class II Θ-SAPs. The subset of these polygons that are Θ_1-SAPs are the polygons that strand exchanges were performed on. From each of these generated samples, the chain number ($i = 1, \ldots, 14$), the polygon length and the after strand exchange knot and link types of each Θ-SAP were recorded so quantities of interest, namely link transition probabilities,

could be estimated.

As an example of how a link transition probability is estimated, consider $\omega_1, \omega_2, \ldots, \omega_{t_0}$ to be a sequence of observed polygons for knot type K, generated from one chain of one replication of the CMC Θ-BFACF algorithm. Then an estimate of $\rho_n^{SE}(K \to L)$ (Equation 2.13) can be obtained from the following ratio: the numerator is the number of n-edge polygons with knot type K that give link type L after a strand exchange divided by the run length t_0, and the denominator is the total number of n-edge Θ_1-SAPs divided by the run length t_0. That is, let:

$$X_n(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is an } n\text{-edge } \Theta_1\text{-SAP that goes from knot type } K \\ & \text{to link type } L \text{ after a strand exchange} \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

and

$$Y_n(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is an } n\text{-edge } \Theta_1\text{-SAP} \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

Then define:

$$\overline{X}_{t_0}(n) = \frac{\sum_{i=1}^{t_0} X_n(\omega_i)}{t_0} \tag{3.7}$$

and

$$\overline{Y}_{t_0}(n) = \frac{\sum_{i=1}^{t_0} Y_n(\omega_i)}{t_0} \tag{3.8}$$

so that the link transition probability, $\rho_n^{SE}(K \to L)$, can be estimated by the ratio:

$$\rho_n = \frac{\frac{\sum_{i=1}^{t_0} X_n(\omega_i)}{t_0}}{\frac{\sum_{i=1}^{t_0} Y_n(\omega_i)}{t_0}} = \frac{\overline{X}_{t_0}(n)}{\overline{Y}_{t_0}(n)}. \tag{3.9}$$

Therefore, the link transition probability estimates are a ratio of two averages over the simulation, so ratio estimation techniques will be needed for statistical analysis.

Normally, for a new MCMC simulation, one analyzes all the data based on an important quantity such as average polygon length, to determine first a τ_{exp}, then a τ_{int} for that quantity. With that complete, and depending on other quantities of interest, other integrated autocorrelation times may be needed, as well as a determination of how much data is 'reliable' for statistical estimation. For the CMC Θ-BFACF algorithm, τ_{exp} and τ_{int} were previously determined, based on average polygon length, to be 5.0 billion [108] and 1.68 billion Θ-BFACF moves in parallel [20] respectively. Because of this, here the description of the analysis techniques is given in a different order to the order in which they would normally be performed. In particular, in the descriptions, it is highlighted how the techniques are used with respect to estimating the link transition probabilities for the LSE model studied in this thesis.

Since link transition probabilities are the quantities of interest to be estimated, some initial analysis must first be completed. The data must be shown to be 'reliable' (*cf.* Section 3.5.1); the time to equilibrium, τ_{exp}, must be estimated (*cf.* Section 3.5.2); proper statistical techniques are needed to correctly compute

the standard errors for ratio estimates (*cf.* Section 3.5.3); the time between essentially independent samples, τ_{int}, must be estimated (*cf.* Section 3.5.4); techniques for analyzing link transition probabilities are needed (*cf.* Section 3.5.5); and finally, fits for the asymptotic behaviour of the link transition probabilities must be obtained (*cf.* Section 3.5.6). Details related to this are discussed in the next subsections, starting with the issue of determining which data is 'reliable'.

3.5.1 Reliable Data, $N_{max}(K)$

Since the simulation is run for a finite amount of time, there is only a finite amount of data collected. Due to this finite nature, the observed proportions of large length polygons may not accurately reflect the corresponding proportions determined from the equilibrium distribution. This is because large length polygons are less probable for low z values, and for high z values the integrated autocorrelation time becomes large. Therefore, it is important to determine where the data becomes 'unreliable'. The following is based on [106].

The desired cut off point separating the reliable from the unreliable data, is determined by estimating the relative standard error for the value of interest, and determining a tolerance level for the deviation from this relative standard error. Here, for example, suppose the quantity of interest is the successful strand exchange probability for n-edge polygon with a particular knot type K, $\rho_n^{SE}(K \to s)$ (Equation 2.14), which in this section will be represented by $\rho_n(K)$ for simplicity. We want to determine the cut off in terms of the polygon length n, denoted $N_{max}(K)$, where these estimated probabilities become statistically unreliable for all polygons with lengths greater than $N_{max}(K)$.

For each replication $r \in \{1, 2, \ldots, 10\}$, let $\widehat{\rho}_n^{(r)}(K)$ denote the estimated proportion of polygons from replication r that are Θ_1-SAPs of length n (*i.e.* the estimate for $\rho_n(K)$). The standard error of this estimate is denoted $\widehat{SE}(\widehat{\rho}_n^{(r)}(K))$. The estimated relative standard error of $\widehat{\rho}_n^{(r)}(K)$ is then defined to be:

$$\widehat{\delta}_n^{(r)}(K) \coloneqq \begin{cases} \frac{\widehat{SE}(\widehat{\rho}_n^{(r)}(K))}{\widehat{\rho}_n^{(r)}(K)}, & \text{if } \widehat{SE}(\widehat{\rho}_n^{(r)}(K)) \neq 0 \\ \infty, & \textit{otherwise.} \end{cases} \tag{3.10}$$

Now we define $\widehat{\delta}^{(r)}(K)$ to be equal to the minimum estimated relative standard error over all n:

$$\widehat{\delta}^{(r)}(K) \coloneqq \min_n \widehat{\delta}_n^{(r)}(K). \tag{3.11}$$

Therefore, $\widehat{\delta}^{(r)}(K)$ is the smallest relative error of the estimated quantity of interest, $\widehat{\rho}_n^{(r)}(K)$, for the generated data. The most accurate overall data will be for values of n where $\widehat{\delta}_n^{(r)}(K)$ is within some tolerance of $\widehat{\delta}^{(r)}(K)$. So this tolerance level, ϵ, is defined as:

$$\epsilon \coloneqq \min_r(\widehat{\delta}^{(r)}(K) + c) \tag{3.12}$$

where $c \times 100\%$ represents the maximum tolerated deviation from $\widehat{\delta}^{(r)}(K)$. Therefore, to ensure that the estimated standard error $\left(\widehat{SE}(\widehat{\rho}_n^{(r)}(K))\right)$ of the point estimate $\left(\widehat{\rho}_n^{(r)}(K)\right)$ is smaller than the estimate itself,

c must be between 0 and 1 and ϵ must be less than 1, *i.e.* $0 < c < 1$ and $\epsilon < 1$. The exact value of c is somewhat arbitrary, but should be chosen so as to minimize the error introduced in calculations. Here, c is chosen to be $c = 0.005$, representing a 0.5% maximum tolerated deviation from $\widehat{\delta}^{(r)}(K)$. This c is similar to that chosen by Szafron [106].

Let the polygon length where the estimated quantity of interest $\widehat{\rho}_n^{(r)}(K)$ first becomes unreliable, be called $N_{max}^{(r)}(K)$, *i.e.* $N_{max}^{(r)}(K)$ is the first polygon length where the estimated relative standard error at length n is at least the tolerance level and, for all larger polygon lengths $\widehat{\delta}_n^{(r)}(K) \geq \epsilon$. In other words, we want $N_{max}(K)^{(r)}$ to be the first polygon length whose estimate is unreliable, so that any estimate for length $n < N_{max}^{(r)}(K)$ is considered reliable, while for $n \geq N_{max}^{(r)}(K)$ it is considered unreliable. The estimate over all replications, $N_{max}(K)$, is then

$$N_{max}(K) \coloneqq \min_r \{N_{max}^{(r)}(K)\}. \tag{3.13}$$

3.5.2 Time to Equilibrium, τ_{exp}

The previous determination of reliable data assumes that the data is from its equilibrium distribution; however, this may not be the case as it takes time for the entire process to reach equilibrium [89, 106]. The exponential autocorrelation time, τ_{exp}, is the number of time steps required for the data to be negligibly affected by a non-equilibrium starting state, and it is used to estimate the time to reach equilibrium. Two methods are covered here for estimating τ_{exp}: a warm-up analysis to measure $\tau_{exp,W}$ [37, 89, 106], and an estimated potential scale reduction method to measure $\tau_{exp,E}$ [37, 43, 44, 106]. The maximum value obtained from these two methods then gives the final estimate of τ_{exp}.

A *warm-up analysis* estimates $\tau_{exp,W}$ by estimating a finite interval $[0, k]$, called the warm-up interval for the Markov chain, which is an interval that contains $\tau_{exp,W}$; meaning $\tau_{exp,W}$ will be at most k. To estimate k, each of 10 independent replications with $t_0 \gg k$ time steps, are started in the same starting state.

Now define $\{X_t^{(r)}, t \in T\}$ to be the Markov chain for replication $r \in \{1, 2, \ldots, 10\}$, where $X_t^{(r)}$ is the state of replication r at time t. Then, given a real-valued function h (for example the average polygon length across all 14 chains of the CMC Θ-BFACF algorithm), define

$$\bar{h}(X_{t_1,t_2}^{(r)}) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} h(X_t^{(r)}) \tag{3.14}$$

to be the average of $h(X_t^{(r)})$ for replication r, where t_1, t_2 are time steps such that $t_1 \leq t_2$. Then, define

$$\bar{h}(X_{t_1,t_2,10}) = \frac{1}{10} \sum_{r=1}^{10} \bar{h}(X_{t_1,t_2}^{(r)}) \tag{3.15}$$

to be the average of $\bar{h}(X_{t_1,t_2}^{(r)})$ over all 10 replications from time step t_1 to time step t_2.

The warm-up interval $[0, k]$ can then be estimated for $1 \leq j \leq t_0$ with the quantities $\bar{h}(X_{0,j,10})$ (the first j-steps average) and $\bar{h}(X_{j,t_0,10})$ (the last j-steps average). This is possible because, provided the Markov chain has converged, there should be a point k^* where the trend of the sequence of the first j-steps average,

$\bar{h}(X_{0,j,10})$, dissipates for all $k^* \leq j \leq t_0$; therefore, k is bounded by k^*, *i.e.* $k \leq k^*$. Similarly, there should be a point k_* where the trend of the sequence of the last j-steps average, $\bar{h}(X_{j,t_0,10})$, dissipates for all $k_* \leq j \leq t_0$, implying $k \leq k_*$. Therefore, from a warm-up analysis, the time it takes for the process to lose the effect of the starting states is estimated to be $k = \max\{k^*, k_*\}$. This can be done for each replication r, giving $\tau_{exp,W}^{(r)} = k^{(r)}$, so for all replications

$$\tau_{exp,W} \coloneqq \max_r \{\tau_{exp,W}^{(r)}\}. \tag{3.16}$$

This warm-up analysis method has two notable limitations. First, it cannot distinguish whether the simulation is sampling from a local or a global distribution. This is problematic because the CMC Θ-BFACF algorithm works on a very local scale, meaning there is little change in a Θ-SAP from step t to $t+1$ which could allow the possibility of the sample being generated from a local equilibrium [89]. Additionally, it is possible that a single non-equilibrium starting state for all replications may take a very long time to dissipate [37, 106]. Due to these limitations the second method called estimated potential scale reduction is also considered for determining τ_{exp}.

In the estimated potential scale reduction method, instead of starting all the replications with the same state, each replication is started from a distinct over-dispersed initial state, $s_0^{(r)}$ for $1 \leq r \leq 10$. Fishman [37] states that the warm-up interval can be found if there exists a $k \leq t_0$, such that for each replication the terms $\left(\bar{h}(X_{0,j}^{(r)}) \text{ for } j \in \{k, k+1, \ldots, t_0\}\right)$ converge to a common region. With each replication having different starting states, the chance that they all converge to the same local equilibrium, as may happen in warm-up analysis, is greatly reduced. A numerical method was developed by Gelman and Rubin [44] to determine if a common region has been reached. The method is referred to as *estimated potential scale reduction* and is presented next.

Each replication r is again run for t_0 time steps, then define

$$B_{10,j} = \frac{j+1}{10-1} \sum_{r=1}^{10} \left(\bar{h}(X_{0,j}^{(r)}) - \bar{h}(X_{0,j,10})\right)^2 \tag{3.17}$$

to be the *between replication variance* and

$$W_{10,j} = \frac{1}{10j} \sum_{r=1}^{10} \sum_{i=0}^{j} \left(\bar{h}(X_i^{(r)}) - \bar{h}(X_{0,j}^{(r)})\right)^2 \tag{3.18}$$

to be the *within replication variance*. Gelman [43] showed that the convergence of the Markov chain to its equilibrium can be detected by monitoring the convergence of the sequence $\sqrt{\widehat{R_j}}$, $j \in 1, \ldots, t_0$ where

$$\sqrt{\widehat{R_j}} \coloneqq \sqrt{\frac{\widehat{\text{var}}(h(X_j))}{W_{10,j}}}. \tag{3.19}$$

This sequence's elements are the estimated potential scale reductions. For $j \leq t_0$, $\sqrt{\widehat{R_j}}$ reduces to:

$$\sqrt{\widehat{R_j}} = \sqrt{\frac{j}{j+1} + \frac{1}{j+1}\frac{B_{10,j}}{W_{10,j}}}. \tag{3.20}$$

As all the replications begin to converge to the equilibrium distribution, *i.e.* as $t_0 \to \infty$, $\sqrt{\widehat{R_j}}$ will go to 1, meaning that one replication is no longer statistically different from any other replication. So, Gelman and Rubin state that if there exists some $k < t_0$, such that $\sqrt{\widehat{R_j}}$ is less than 1.1 for all $k \leq j \leq t_0$ then the simulation has converged [44]. Another way to look at this is that $\sqrt{\widehat{R_j}}$ is a measure of how large the between replication variance is relative to the within replication variance, so $\sqrt{\widehat{R_j}} \leq 1.1$ means that the between replication standard deviation is less than 10% larger than the within replication standard deviation. This observed value for k is then the upper limit of the warm-up interval for all the replications and we set

$$\tau_{exp,E} \coloneqq k. \tag{3.21}$$

Taking both methods for estimating τ_{exp} into account, τ_{exp} is then estimated to be:

$$\tau_{exp} \coloneqq \max\{\tau_{exp,W}, \tau_{exp,E}\}. \tag{3.22}$$

Another name for τ_{exp} is the *burn-time* of the simulation. Because data sampled prior to the time τ_{exp} may not be from the equilibrium distribution, this data can be discarded in order to minimize the chance of any bias from non-equilibrium data. However, if τ_{exp} is less than 5% of the total simulation run, then the statistical error introduced by including data collected prior to τ_{exp} is minimal, so data before τ_{exp} can be included in any estimate with little negative effect [89].

3.5.3 Ratio Estimation

Once the reliable data cut off point and the time to equilibrium have been established, we can now begin to look at estimating the link transition probabilities of Equation 2.13:

$$\rho_n^{SE}(K \to L) = \frac{p_n^{\Theta_{SE}}(K \to L)}{p_n^{\Theta_1}(K)}.$$

As established earlier, link transition probability estimates are the ratio of two averages (Equation 3.9), therefore we are interested in ratio estimation and so must use a specialized technique. The following is based on [104, 106].

Suppose $\{(X_i), i = 1, \ldots, t_0\}$ and $\{(Y_i), i = 1, \ldots, t_0\}$ are two sequences of independent and identically distributed random variables with positive means, μ_X and μ_Y, and variances σ_X^2 and σ_Y^2. Formally, this is: $\mu_X \coloneqq E[X_i] \neq 0$, $\mu_Y \coloneqq E[Y_i]$, $\sigma_X^2 \coloneqq E[(X_i - \mu_X)^2] < \infty$, $\sigma_Y^2 \coloneqq E[(Y_i - \mu_Y)^2] < \infty$, and $\sigma_{X,Y}^2 \coloneqq E[(X_i - \mu_X)(Y_i - \mu_Y)] < \infty$ for $i = 1, \ldots, t_0$. Then define, $R \coloneqq \frac{\mu_X}{\mu_Y}$ to be the desired ratio, and

$$\overline{R}_{t_0} \coloneqq \begin{cases} \frac{\overline{X}_{t_0}}{\overline{Y}_{t_0}}, & \text{if } \overline{Y}_{t_0} \neq 0 \\ 0, & \textit{otherwise.} \end{cases} \tag{3.23}$$

For example, $\overline{X}_{t_0}$ and $\overline{Y}_{t_0}$ could correspond to $\overline{X_{t_0}}(n)$ and $\overline{Y_{t_0}}(n)$, respectively, in Equations 3.7 and 3.8 if $\omega_1, \omega_2, \ldots, \omega_{t_0}$ were an independent sample of polygons from the equilibrium distribution, and then the desired ratio R is $\rho_n^{SE}(K \to L)$.

Fishman [37] proved that $\overline{R}$ is a biased estimator of R, with the dominant term of the bias being:

$$\frac{R_{t_0}}{t_0}\left[\frac{\sigma_Y^2}{\mu_Y^2} - \frac{\sigma_{X,Y}^2}{\mu_X \mu_Y}\right].$$

In order to reduce the bias, Fishman recommends using the estimator

$$\widetilde{R} \coloneqq \overline{R}_{t_0}\left[1 + \frac{1}{t_0}\left(\frac{\widehat{\sigma}_{X,Y}^2}{\overline{X}_{t_0}\overline{Y}_{t_0}} - \frac{\widehat{\sigma}_Y^2}{\overline{Y}_{t_0}^2}\right)\right], \tag{3.24}$$

where $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_Y^2$ are the sample variances for X and Y, and $\widehat{\sigma}_{Y,X}^2$ is the sample covariance of X and Y, formally:

$$\widehat{\sigma}_X^2 \coloneqq \frac{1}{t_0 - 1}\sum_{i=1}^{t_0}(X_i - \overline{X}_{t_0})^2,$$

$$\widehat{\sigma}_Y^2 \coloneqq \frac{1}{t_0 - 1}\sum_{i=1}^{t_0}(Y_i - \overline{Y}_{t_0})^2,$$

$$\widehat{\sigma}_{X,Y}^2 \coloneqq \frac{1}{t_0 - 1}\sum_{i=1}^{t_0}(X_i - \overline{X}_{t_0})(Y_i - \overline{Y}_T).$$

Tin [109] then showed $\widetilde{R}$ and its associated variance,

$$\widehat{\mathrm{var}}(\widetilde{R}) \coloneqq \frac{\widetilde{R}^2}{t_0}\left[\frac{\widehat{\sigma}_X^2}{\overline{X}_{t_0}^2} + \frac{\widetilde{\sigma}_Y^2}{\overline{Y}_{t_0}^2} - \frac{2\widehat{\sigma}_{X,Y}^2}{\overline{X}_{t_0}\overline{Y}_{t_0}}\right] \tag{3.25}$$

can be considered essentially unbiased estimators of R and $\widehat{\mathrm{var}}(R)$ for large enough t_0.

Using these unbiased estimators, the ratio R can be estimated, along with a 95% confidence interval (see [106] for full details about this). Note, however, that there is an assumption in this analysis that, for example, the (X_i)'s are independent. In order to use these formulae to estimate the link transition probabilities we need to determine a way to obtain 'independent' data from the CMC Θ-BFACF algorithm. This is reviewed next.

3.5.4 Time to Independence, $\tau_{int,A}$

The ratio estimation technique just discussed can be used to estimate link transition probabilities if one has independent data. However, because the data generated by the CMC Θ-BFACF algorithm is not independent, in order to use this ratio estimation technique, the data must first be separated into 'independent' *blocks*. Two states are said to be *essentially independent* if the number of time steps between them is deemed sufficiently large *i.e.* when the correlation between the two states is essentially zero. So, the greater the amount of time between two states, the less correlated the states will be. The acceptable time is calculated using the autocorrelation function, $\alpha_A(h)$, of an observable A [89]. Formally, given a stochastic process $\{A(X_t), t \in T\}$, two variables $A(X_{t_1})$ and $A(X_{t_2})$ are essentially independent if:

$$|t_2 - t_1| \geq 2\tau_{int,A} \tag{3.26}$$

where, assuming the chain is in equilibrium, $\tau_{int,A}$ is the integrated autocorrelation time,

$$\tau_{int,A} = \frac{1}{2}\sum_{h=-\infty}^{\infty} \alpha_A(h) = \frac{1}{2}\sum_{h=-\infty}^{\infty} \frac{\mathrm{Cov}(A(X_0), A(X_h))}{\mathrm{var}(A(X_0))}. \tag{3.27}$$

The number of total simulation time steps, t_0, is then divided into blocks of length $2\tau_{int,A}$ time steps to ensure all blocks, and so data points in different blocks, are independent of each other and the ratio estimation technique can then be used.

For example, for a particular quantity of interest, A, the autocorrelation function, $\tau_{int,A}$, is used to determine the first time $t_2 > t_1$ where the autocorrelation between the two quantities is insignificant at a desired confidence level, therefore giving two time steps, t_2 and t_1, which are essentially independent. For the full details and calculation of $\tau_{int,A}$ see [106].

Using the example from Equations 3.5-3.8, with $\tau = \max\{\tau_{int,X_n}, \tau_{int,Y_n}\}$, and $m = \frac{t_0}{2\tau}$ blocks, define:

$$\overline{X}_{t_1,t_2}(n) = \frac{1}{t_2 - t_1}\sum_{t=t_1}^{t_2-1} X(\omega_t) \tag{3.28}$$

and

$$\overline{Y}_{t_1,t_2}(n) = \frac{1}{t_2 - t_1}\sum_{t=t_2}^{t_2-1} Y(\omega_t). \tag{3.29}$$

Then $\{(\overline{X}_{2\tau j,2\tau(j+1)}), j = 0,\ldots,m-1\}$ and $\{(\overline{Y}_{2\tau j,2\tau(j+1)}), j = 0,\ldots,m-1\}$ are the two sequences of independent and identically distributed variables needed for the ratio estimation of $\rho_n^{SE}(K \to L)$ by ρ_n from Equation 3.9.

It has been shown for the Θ structure that polygons 1.68×10^9 Θ-BFACF moves in parallel apart are essentially independent, and so subdividing the sampled data into blocks of 1.68×10^5 consecutive data points form essentially independent blocks of data [20].

Thus, this method gives a way to estimate $\rho_n^{SE}(K \to L)$ for a given polygon length n, however, we are also interested in exploring the behaviour of these link transition probabilities as a function of n. All these estimates come from the same CMC Θ-BFACF simulation run, so they will be correlated. In order to take this correlation between different polygon lengths into account in the analysis, further methods must be used and they are discussed next.

3.5.5 Analyzing Link Transition Probability as a Function of Polygon Length

As the previous section stated, two estimates of $\rho_n^{SE}(K \to L)$ for $n = n_1$ and $n = n_2$ are correlated, therefore this correlation must be accounted for. There are two methods for estimating link transition probabilities in order to take this correlation into account: the fixed-n method and the grouped-n method. Each of these methods use ratio estimation and will be review next based on [106].

The goal is to obtain estimates of $\rho_n^{SE}(K \to L)$ for a sequence of polygon lengths $n_1, n_2, \ldots, N_{max}(K)$ and to use these to determine the asymptotic behaviour of $\rho_n^{SE}(K \to L)$ to study the conjectures in Equations 2.18-2.21. For example, for *Case 1* we would like to estimate parameters such as m and $t_{K,L}^{SE}$ in the fitting

form seen in the next section in Equation 3.30. As discussed in the last section (*cf.* Equation 3.9), we can obtain ratio estimates $\{\rho_n, 14 \leq n \leq N_{max}(K)\}$ for $\{\rho_n^{SE}(K \to L),\ 14 \leq n \leq N_{max}(K)\}$, however, these are correlated.

For the fixed-n method, instead of using all these estimates, an essentially independent subsample is used. This subsample is obtained using an appropriate integrated autocorrelation time, $\tau = \tau_{int,\rho}$, and is determined by analyzing the autocorrelation of the ρ_n's. The subsample is then $\{\rho_{14+2\tau k}, k \geq 0\}$. With this subsample, estimates for parameters such as m and $t_{K,L}^{SE}$ can be obtained using weighted least-squares regression and fitting a curve to the data (covered in the next section).

For the grouped-n method, all the estimates for all the lengths from 14 to $N_{max}(K)$ are used; however, they are now grouped into essentially independent blocks of data according to polygon length. This method is done, for example, by grouping together the $\overline{X}_{t_0}(n)$'s, from $n = n_1$ to $n_2 - 2$, to obtain $\widetilde{X}_{t_0}(n) = \sum_{n=n_1}^{n_2-2} \overline{X}_{t_0}(n)$; this becomes a blocked estimate. Similarly group the $\overline{Y}_{t_0}(n)$'s. From these blocked estimates a grouped-n estimate of $\rho_{n_1}^{SE}(K \to L)$ for the first block is obtained. The rest of the data is similarly blocked using the block size $(n_2 - n_1)$, to obtain estimates of $\rho_{n_j}^{SE}(K \to L), n_j \leq N_{max}(K)$, where $n_{j+1} - n_j = n_2 - n_1$. When full analysis is performed (see [106]) it can be shown that the grouped-n estimates have the same asymptotic behaviour as the fixed-n estimates as n tends to infinity. In general, the grouped-n approach provides a method for using all the Θ-SAP data from length n_1 to $n_2 - 2$, in order to estimate the desired ratio as a function of length [89]. The width of the interval $(n_2 - n_1)$ is determined using a procedure similar to the method for determining the subsample for the fixed-n method, *i.e.* the block size is chosen to ensure the blocked estimates are essentially independent of each other. Therefore, the grouped-n method can be used to estimate the link transition probabilities and then estimates for parameters, such as m and $t_{K,L}^{SE}$, can be obtained using weighted least-squares regression and curve fitting.

For simplicity in the above discussion, we have used an example that only involves polygons generated from one chain of one replication of the CMC Θ-BFACF algorithm. In practice, averages are calculated across all 14 chains and over all 10 replications to obtain link transition probabilities. The algorithms used for this all take this into account correctly [106].

3.5.6 Fitting Data

Since the grouped-n method uses more of the simulation data, it is the method primarily used in this work. In order to determine the asymptotic behaviour of $\rho_n^{SE}(K \to L)$ to study the conjectures in Equations 2.18-2.21, fits for the grouped-n estimates are generated using weighted non-linear regression functions in the statistical package R. From these fits the asymptotic values of $t_{K,L}^{SE}$ for *Case 1* and $C_{K,L}^{SE}$ for *Case 2* as $n \to \infty$ can be estimated (see Section 2.3 for details on these asymptotic values).

From Equation 2.10, in *Case 1* we expect as $n \to \infty$ that the link transition probability should go to zero. In order to fit this, the following three forms are used:

For *Case 1*:

$$\rho_n^{SE}(K \to L) \sim mn^{-t_{K,L}^{SE}} \tag{3.30}$$

indicated with a $^{(1)}$

$$\rho_n^{SE}(K \to L) \sim mn^{-t_{K,L}^{SE}} + cn^{-t_{K,L}^{SE}-1} \tag{3.31}$$

indicated with a $^{(2)}$

$$\rho_n^{SE}(K \to L) \sim mn^{-t_{K,L}^{SE}} + cn^{-2t_{K,L}^{SE}} \tag{3.32}$$

indicated with a $^{(3)}$,
where if ρ_n^{SE} is indeed going to zero, the estimate for $t_{K,L}^{SE}$ should be 1. The form of the fit used will be indicated with the corresponding superscript next to the $t_{K,L}^{SE}$ estimate.

From Equation 2.11, in *Case 2* we expect as $n \to \infty$ the link transition probability should go to a nonzero positive constant. However, here $n \ll \infty$ to remain within the reliable data cut off, so *Case 1* will still play a significant role in the estimates at these small values of n. Therefore we must eliminate *Case 1*'s influence to get a true estimate for *Case 2*. *Case 1*'s influence is eliminated by subtracting the set of links resulting from *Case 1*, denoted by $\Gamma_{SE,C1}$, from the full set of links, $\Psi_{SE,L}$. So the new desired probability becomes:

$$\rho_n^{SE*}(K \to L) \coloneqq \frac{p_n^{SE}(K \to L)}{\sum\limits_{\gamma \in \Psi_{SE,L} - \Gamma_{SE,C1}} p_n^{SE}(K \to \gamma)} = \frac{\rho_n^{\Theta_1}(K \to L)}{\rho_n^{SE}(K \to \Psi_{SE,L} - \Gamma_{SE,C1})}. \tag{3.33}$$

It is expected that $\rho_n^{SE*}(K \to L)$ will have the same behaviour as $\rho_n^{SE}(K \to L)$ when $n \to \infty$, so $\rho_n^{SE*}(K \to L) \sim \rho_n^{SE}(K \to L) \sim C_{K,L}^{SE} > 0$. Therefore, to fit this new $\rho_n^{SE*}(K \to L)$ the following three forms are used:
For *Case 2*:

$$\rho_n^{SE*}(K \to L) \sim b + mn^{-t_{K,L}^{SE}} \tag{3.34}$$

indicated with a $^{(4)}$

$$\rho_n^{SE*}(K \to L) \sim b + mn^{-t_{K,L}^{SE}} + cn^{-t_{K,L}^{SE}-1} \tag{3.35}$$

indicated with a $^{(5)}$

$$\rho_n^{SE*}(K \to L) \sim b + mn^{-t_{K,L}^{SE}} + cn^{-2t_{K,L}^{SE}} \tag{3.36}$$

indicated with a $^{(6)}$,
where the estimate for b will be $C_{K,L}^{SE}$ as $n \to \infty$. Again, the form of the fit used will be indicated with the corresponding superscript next to the $C_{K,L}^{SE}$ estimate.

The quality of these fits is then determined using a χ^2 test for goodness of fit. The following is based on [89]. Let $\{(N_1, R_1), \ldots, (N_m, R_m)\}$ represent the observed data points that are being fit (N for length

n, R for ratio). Then, $\left\{\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_m^2\right\}$ represents the estimated variances of the estimates $(R_1, \ldots, R_m)$. Let $\{(N_1, F_1), \ldots, (N_m, F_m)\}$ be the expected data points of the fit at points $(N_1, \ldots, N_m)$ (F for fits). The goodness of fit (G.O.F.) test statistic is then defined as:

$$\Lambda = \sum_{i=1}^{m} \frac{(R_i - F_i)^2}{\widehat{\sigma}_i^2}. \tag{3.37}$$

If there are M parameters in the model being fit, then there are $m - M$ degrees of freedom (df) for the test. The value of Λ can then be compared to the χ^2 distribution with $m - M$ degrees of freedom, and a p value can be obtained when the probability of Λ exceeds that of the χ^2 distribution. A p value less than 0.05 for a fit indicates that the fit may not be appropriate.

3.6 Chapter 3 Summary

This chapter discussed the algorithms used to generate Θ-SAPs and the analysis techniques used to properly analyze this generated data. The chapter began by introducing definitions for Markov chains and Monte Carlo simulations, which ultimately lead to the introduction of the main algorithm for data generation, the CMC Θ-BFACF algorithm. This algorithm is ergodic on the set of all class II Θ-SAPs with knot type K, meaning starting with a class II Θ-SAP, ω with knot type K, any other class II Θ-SAP ω' with knot type K can be reached. Simulation specifics were then given, including that the result of the simulation was $560,000,000$ sampled Θ-SAPs for each of the starting knot types $K = \phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+$, and 5_2^-.

With the data generated, analysis techniques were then reviewed to explain how reliable, essentially independent data within the equilibrium distribution can be obtained for analysis. Since link transition probabilities are estimated by a ratio, ratio estimate techniques were covered. Then issues related to the study of how the link transition probabilities behave as a function of polygon length n were discussed, resulting in the grouped-n method being the chosen approach for addressing these issues. Finally the forms of the fits for both *Case 1* and *Case 2* link transition probabilities were introduced, with the χ^2 test for goodness of fit established for determining the quality of the fit.

The next chapter will begin with the results from the analysis techniques described here, and then will present results from the LSE model.

Chapter 4

Analysis and Link Transition Probability Estimates for the LSE Model

Chapter 2 introduced a new Local Strand Exchange (LSE) model in order to model recombinase-DNA interactions, and Chapter 3 described the algorithms used to generate the data and the analysis techniques used to analyze the generated data. This chapter will begin with the results from the analysis techniques outlined in Chapter 3 and then will present the results from the LSE model. Regarding these results, first the grouped-n estimates of the successful strand exchange probability will be examined, and then results for link transition probability estimates will be presented in subsequent subsections. In these subsequent subsections, first the DLE results and then the IKE results will be presented, each taking both *Case 1* and *Case 2* into consideration. The section closes with a brief summary and discussion. The chapter will end with a very brief comparison between the LSE model results and those from the Local Strand Passage (LSP) model (the topoisomerase-DNA interaction model).

It will be shown that the link transition probability estimates resulting from studying the LSE model support the conjectures presented in Equations 2.18-2.21, with the link transition probability estimates in *Case 1* going to zero as polygon length goes to infinity, and in *Case 2* going to a nonzero constant depending on the resulting link type. It will also be shown that there are similar trends between the LSE and LSP models.

4.1 Analysis Results

In this section the techniques reviewed in Section 3.5 will be applied to the CMC Θ-BFACF simulation data of the LSE model and the associated results will be presented. Specifically: first $N_{max}(K)$ will be determined for each starting knot type K (*cf.* Section 4.1.1); next τ_{exp} will be determined for knot type $K = 5_1^-$, as an example (*cf.* Section 4.1.2); and finally the grouped-n block size will be determined for several link transition probability estimates (*cf.* Section 4.1.3).

	Replication r										
K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Overall**
ϕ	2056	2088	2076	2080	2044	2086	2052	2086	2088	2084	**2044**
3_1^+	2342	2340	2328	2290	2306	2270	2320	2306	2214	2234	**2214**
3_1^-	2126	2144	2186	2144	2148	2136	2208	2166	2164	2198	**2126**
4_1	2374	2418	2368	2396	2438	2372	2446	2342	2456	2364	**2342**
5_1^+	2510	2658	2610	2652	2654	2682	2664	2686	2530	2618	**2510**
5_1^-	2474	2514	2506	2530	2482	2442	2450	2472	2496	2458	**2442**
5_2^+	2618	2616	2672	2666	2536	2658	2596	2634	2688	2630	**2536**
5_2^-	2518	2566	2508	2552	2526	2574	2542	2572	2640	2568	**2508**

Table 4.1: $N_{max}^{(r)}(K)$ values for each replication $r \in \{1, 2, \ldots, 10\}$ for all knot types $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^-, 5_1^+, 5_2^-, \text{ and } 5_2^-\}$. The last column corresponds to the overall maximum polygon length, $N_{max}(K)$.

4.1.1 Estimating $N_{max}(K)$

In order to ensure only reliable data is analyzed the maximum polygon length, $N_{max}(K)$, defined in Section 3.5.1, must be estimated. Using the method outlined in Section 3.5.1, $N_{max}^{(r)}(K)$ was estimated for each starting knot type K; and the estimates are shown in Table 4.1. The general procedure used to determine the estimates will be presented in detail for $K = \phi$ as an example.

The relative standard errors (Equation 3.10) for the successful strand exchange probability, $\rho_n^{SE}(\phi \to s)$, as estimated from each replication, are presented in Figure 4.1. The lower dashed line of this graph shows the minimum relative standard error over all 10 replications, $\min_r\{\delta_n^{(r)}(\phi)\}$ (Equation 3.11), which here is 0.00013. The upper dashed line is the maximum tolerated deviance, ϵ (Equation 3.12), from $\min_r\{\delta_n^{(r)}(\phi)\}$, which here is 0.00513. Therefore, $N_{max}(\phi)$ is the first value, n, where all $\min_r\{\delta_n^{(r)}(\phi)\}$ exceed ϵ_*. In this example over all replications $N_{max}(\phi) = 2044$, so only unknotted polygons with lengths less than 2044 are considered reliable.

See Appendix A.1 for the corresponding plots for all other knot types.

Now that $N_{max}(K)$ has been estimated, the next step is to estimate the time to equilibrium via τ_{exp}.

4.1.2 Estimating τ_{exp}

In Section 3.5.2 two methods were described for estimating τ_{exp}: a warm-up analysis to obtain $\tau_{exp,W}$, and the estimated potential scale reduction method to obtain $\tau_{exp,E}$. These analyses were performed on all eight starting knot types for the LSP by Szafron [106], with estimates supporting that τ_{exp} is less than 5% of each simulation's total length.

Here, as an example, the complete results from both of these estimation methods will be shown for starting

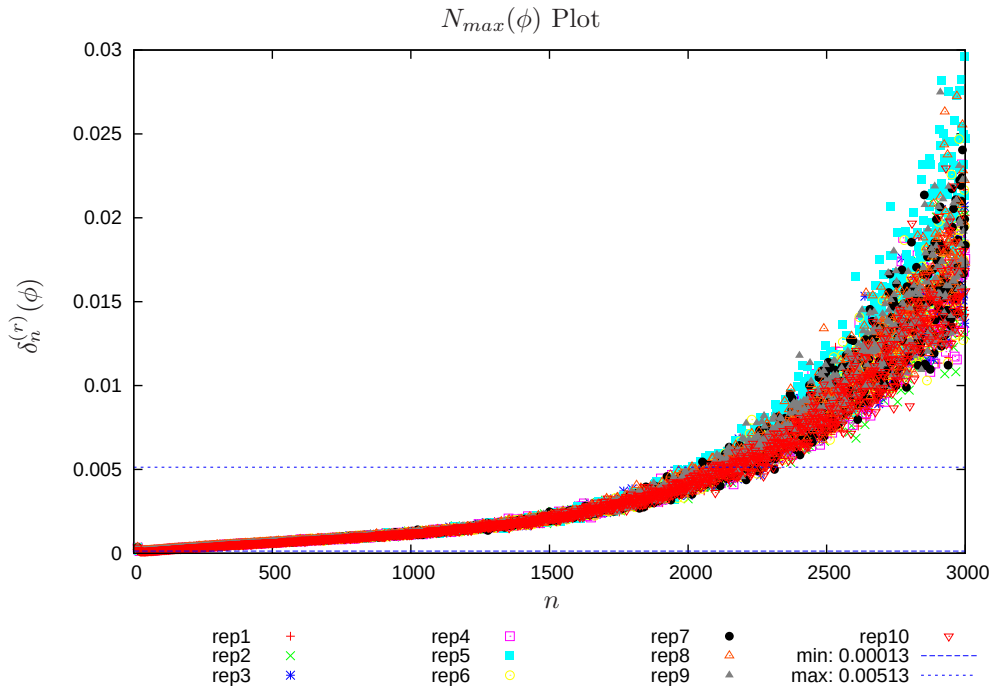


Figure 4.1: Plot of the relative standard error starting with ϕ to determine $N_{max}(\phi)$.

knot type $K = 5_1^-$. This analysis is presented for $K = 5_1^-$ for the LSE model, because for $K = 5_1^-$, the realized CMC Θ-BFACF data displayed behaviour for one billion time-steps in one replication (at 160 to 161 billion Θ-BFACF moves in parallel in replication six) which did not follow the behaviour of the other data generated for this replication. It appeared that a local equilibrium may have been entered, so to ensure that the entire simulation reached the global equilibrium, τ_{exp} was estimated.

In order to determine $\tau_{exp,W}$, the real-valued function h used was the average polygon length across all 14 chains, $\left(h(X_j^{(1)}), j = 0, \ldots, t_0\right)$. This function was chosen as Szafron [106] showed that, of all the functions of interest, it was the function with the largest variance. Therefore, define $\bar{h}(X_{t_1,t_2}^{(r)})$ to be the average polygon length across all 14 chains for replication r from time step t_1 to t_2, and $\bar{h}(X_{t_1,t_2,10})$ to be the average polygon length across all 14 chains over all 10 replications from time step t_1 to t_2.

Figure 4.2 illustrates the average length of the polygons across the 14 chains $\left(h(X_j^{(1)}), j = 0, \ldots, t_0\right)$, the first j-steps averages $\left(\bar{h}(X_{0,j}^{(1)}), j = 0, \ldots, t_0\right)$ and the last j-step averages $\left(\bar{h}(X_{j,t_0}^{(1)}), j = 0, \ldots, t_0\right)$ for replication one from time step 0 to $t_0 = 400$ billion Θ-BFACF moves in parallel. From this plot, k^* is chosen to be when the first j-steps averages minimally fluctuate as j increases, and k_* is chosen to be when the last j-steps averages minimally fluctuate as j increases. This gives $k = \max\{k^*, k_*\}$ for replication $r = 1$. In this example, from Figure 4.2, both k^* and k_* are about 10 billion Θ-BFACF moves in parallel, meaning that after 10 billion Θ-BFACF moves the effect of the starting states has dissipated. This analysis was done for each replication resulting in the overall $\tau_{exp,W}$ (Equation 3.16). Plots for the other nine replications can be found in Appendix A.2, and a summary table for all 10 replications is presented in Table 4.2. From Table

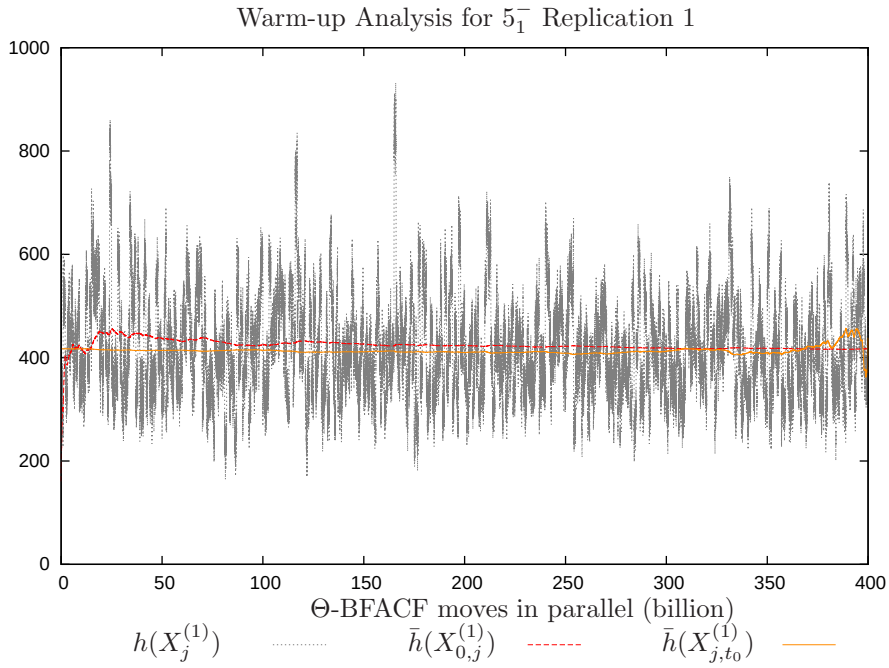


Figure 4.2: Plot of warm-up analysis for replication one. $\left(h(X_j^{(1)}), j = 0, \ldots, t_0\right)$ are the average polygon lengths over all 14 chains at time step j. $\left(\bar{h}(X_{0,j}^{(1)}), j = 0, \ldots, t_0\right)$ are the first j-steps averages, and $\left(\bar{h}(X_{j,t_0}^{(1)}), j = 0, \ldots, t_0\right)$ are the last j-steps averages. Each is plotted over all 400 billion Θ-BFACF moves in parallel.

4.2, by Equation 3.16, $\tau_{exp,W}$ = 12 billion Θ-BFACF moves in parallel.

To address some limitations of the warm-up analysis, estimated potential scale reductions analysis was also performed. For this method Fishman [37] states that if the $\left(\bar{h}(X_{0,j}^{(r)}), j = 0, \ldots, t_0\right)$ for each replication overlap into a common region then there exists a k which is an estimate for τ_{exp} (see Section 3.5.2 for more details). Figure 4.3 illustrates the sequences $\left(\bar{h}(X_{0,j}^{(r)}), j = 0, \ldots, t_0\right)$ for all 10 replications. From this plot, it can be seen that all replications appear to converge to a common region. To quantitatively confirm this observation, Figure 4.4 shows a plot of the estimated potential scale reductions $\sqrt{\widehat{R_j}}$. From Section 3.5.2, it is known that as the between replication variance approaches the within replication variance, $\sqrt{\widehat{R_j}}$ should go to 1, as is the case in Figure 4.4. To get a better look, the first 15 billion Θ-BFACF moves in parallel are zoomed in on in Figure 4.5. Using the value of 1.1, or 10%, as a cut off for $\sqrt{\widehat{R_j}}$, it can be seen that after about 1 billion Θ-BFACF moves in parallel the simulation appears to have converged to its equilibrium distribution. However, to be more conservative, the dashed line representing 1.05, or 5.0%, is used as the cut off value. Using this more conservative criteria for convergence to the equilibrium distribution, from Figure 4.5, $\tau_{exp,E}$ is approximately 5 billion Θ-BFACF moves in parallel.

Finally, from Equation 3.22, the estimate for τ_{exp} is the maximum of $\tau_{exp,W}$ and $\tau_{exp,E}$, so therefore, τ_{exp} = 12 billion Θ-BFACF moves in parallel. This means that, after 12 billion Θ-BFACF moves in parallel, the effect of the starting states has dissipated, and the equilibrium distribution has been reached. As discussed in

Replication	k^*	k_*	k
1	10	10	10
2	4	8	8
3	10	10	10
4	10	12	12
5	10	10	10
6	12	10	12
7	8	8	8
8	5	10	10
9	5	10	10
10	5	12	12
Overall			**12**

Table 4.2: Warm-up analysis estimates for all 10 replications, $r \in \{1, 2, \ldots, 10\}$, for knot-type $K = 5_1^-$.

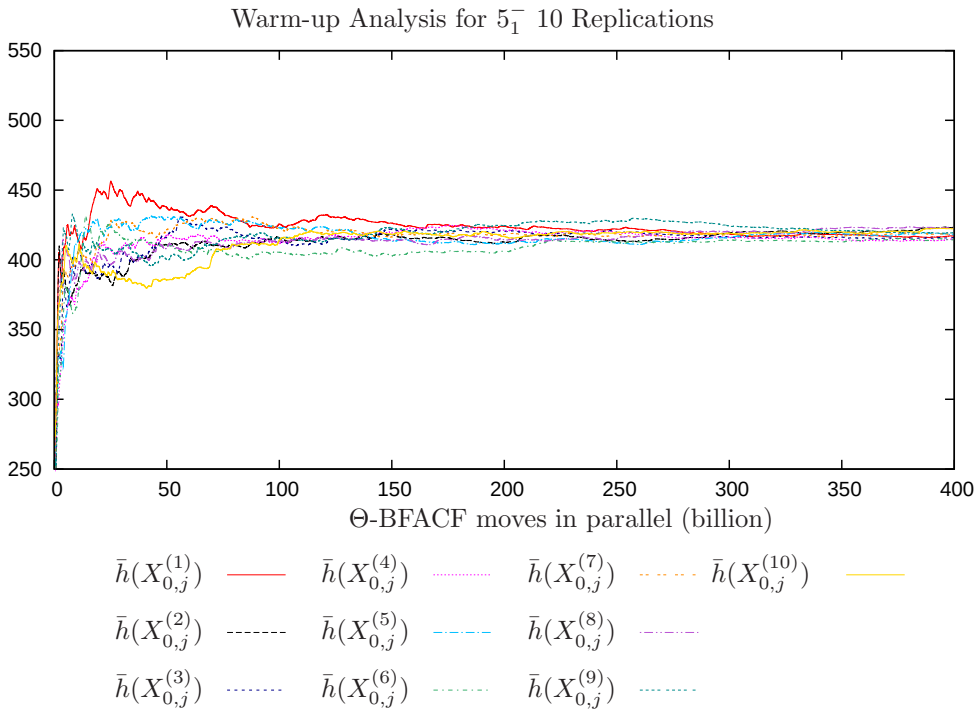


Figure 4.3: Plot of $\left(\bar{h}(X_{0,j}^{(r)}), j = 0, \ldots, t_0\right)$ for all 10 replications for knot type 5_1^-.

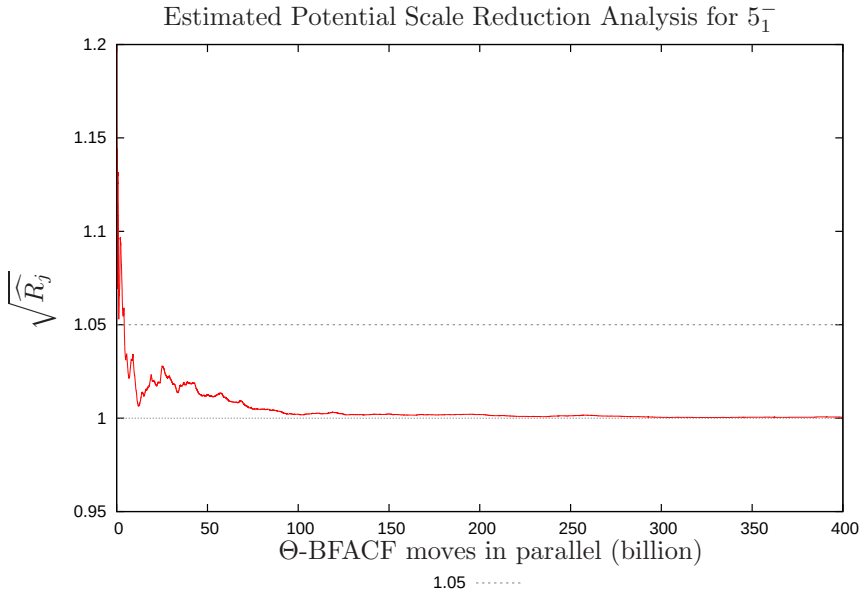


Figure 4.4: Plot of $\sqrt{\widehat{R_j}}$. The dashed line at 1.05 represents a 5.0% cut off value.

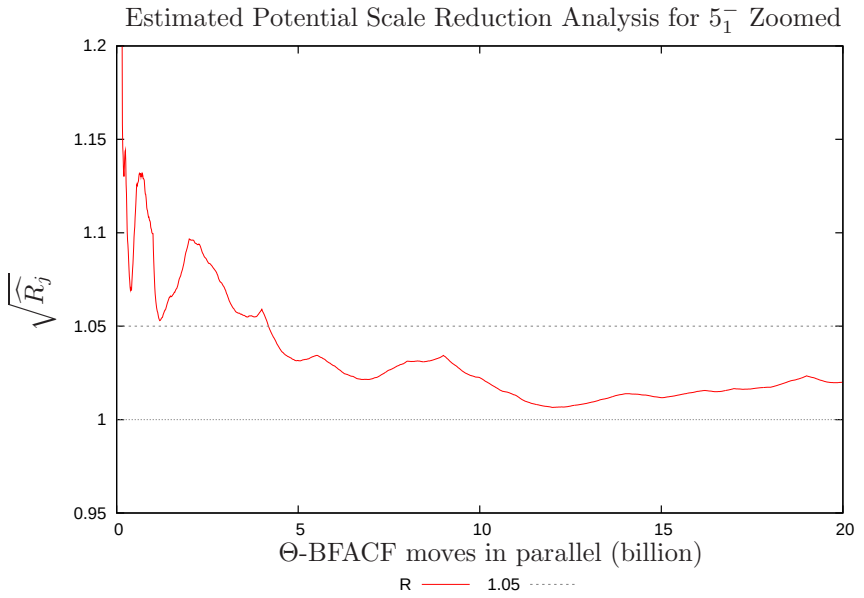


Figure 4.5: Zoomed in plot of $\sqrt{\widehat{R_j}}$. The dashed line at 1.05 represents a 5.0% cut off value.

Section 3.5.2, since τ_{exp} is less than 5% of the total 400 billion Θ-BFACF moves in parallel, the first 12 billion Θ-BFACF moves do not need to be discarded in any subsequent analysis. This analysis provides evidence that the simulation for $K = 5_1^-$ is indeed in the global equilibrium and the possible local equilibrium in replication six should not adversely affect any subsequent analysis.

4.1.3 Estimating a Grouped-n Block Size

To estimate the link transition probabilities using as much data as possible, the grouped-n ratio estimation method will be used (see Section 3.5.5). To use this method, the block size for each different link transition probability must be determined. This block size is determined from plots of the autocorrelation functions for the numerator and denominator of each link transition probability. For polygon lengths $n_2 > n_1$, when the partial autocorrelation between estimates ρ_{n_1} and ρ_{n_2} values is insignificant at a 95% confidence level, the estimates are considered essentially independent. When estimates at two lengths are essentially independent, these two estimates can be in separate grouped-n blocks.

As an example, a complete example of this method is provided for determining the block size needed to estimate $\rho_n^{IKE}(\phi \to \phi)$ as a function of n. The plots required for this are presented in Figure 4.6. The top plot in this figure displays the probability of interest, the estimated link transition probability for going from knot type $K = \phi$ to knot type $K' = \phi$ after a strand exchange, versus polygon length n. The middle row of the figure are plots on two different scales of the partial autocorrelation function (autocorrelation of estimates h n's apart) for the numerator of the link transition probability's estimator: the number of n-edge polygons that go from $K = \phi \to K' = \phi$, versus the associated lag (h). The red line represents the estimated upper boundary of a 95% confidence interval. When the estimated partial autocorrelations plotted are below the red line, the x-axis lag gives the number of acceptable fixed-n samples which can be in the same block. The last row of the figure shows plots on two different scales of the partial autocorrelation function for the denominator of the link transition probability's estimator: the number of n-edge polygons that have a successful SE. The green line represents the estimated upper boundary of the 95% confidence interval. Again, when the estimated partial autocorrelations plotted are below this green line the x-axis gives the number of acceptable fixed-n samples which can be in the same block. The greater of the two lag values is multiplied by two yielding how many n's apart two estimates of $\rho_n^{IKE}(\phi \to \phi)$ must be in order to be considered essentially independent, and hence is an estimate for the grouped-n block size. In Figure 4.6, the largest lag is 50, and so gives a block size of 100, meaning the estimates for $\rho_n^{IKE}(\phi \to \phi)$ with $n = n_1$ and $n = (n_1$+block size) will be in separate blocks. Therefore, for $\rho_n^{IKE}(\phi \to \phi)$ the first block contains polygons with lengths $[14, 112]$ (14 is the smallest Θ-SAP possible), and blocks continue like this until $N_{max}(\phi)$ is reached.

The grouped-n block sizes, using $N_{max}(K)$, for several observed link transitions, $K \to L$, can be seen in in Appendix B in column five of Tables B.1, B.2, B.3, B.4, B.5, B.6, B.7, and B.8.

Now that the reliable data, $N_{max}(K)$, the time to equilibrium, τ_{exp}, and the proper grouped-n block sizes have been estimated the results from the LSE model will be studied.

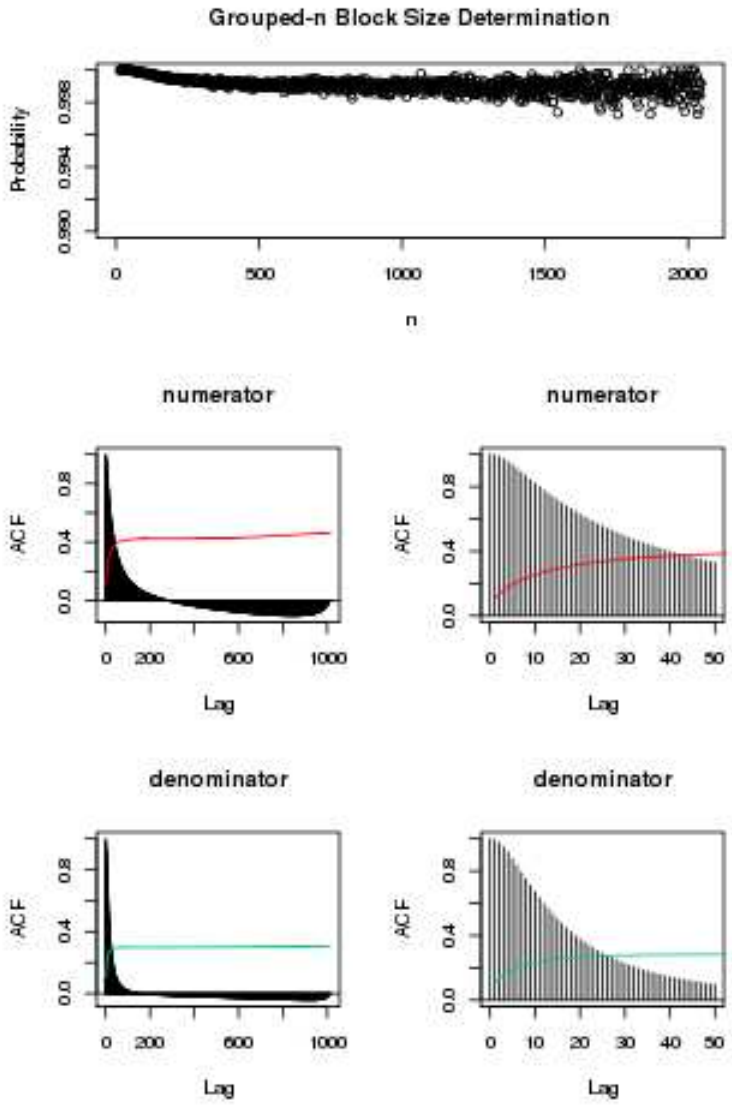


Figure 4.6: A graph for the determination of the grouped-n block size for $\rho_n^{IKE}(\phi \to \phi)$. The top plot is the link transition probability, the middle row is the autocorrelation plots for the numerator, and the bottom row is the autocorrelation plots for the denominator. The block size equals the two times the largest lag when the coloured lines are above the bars representing the estimated partial autocorrelations for each lag.

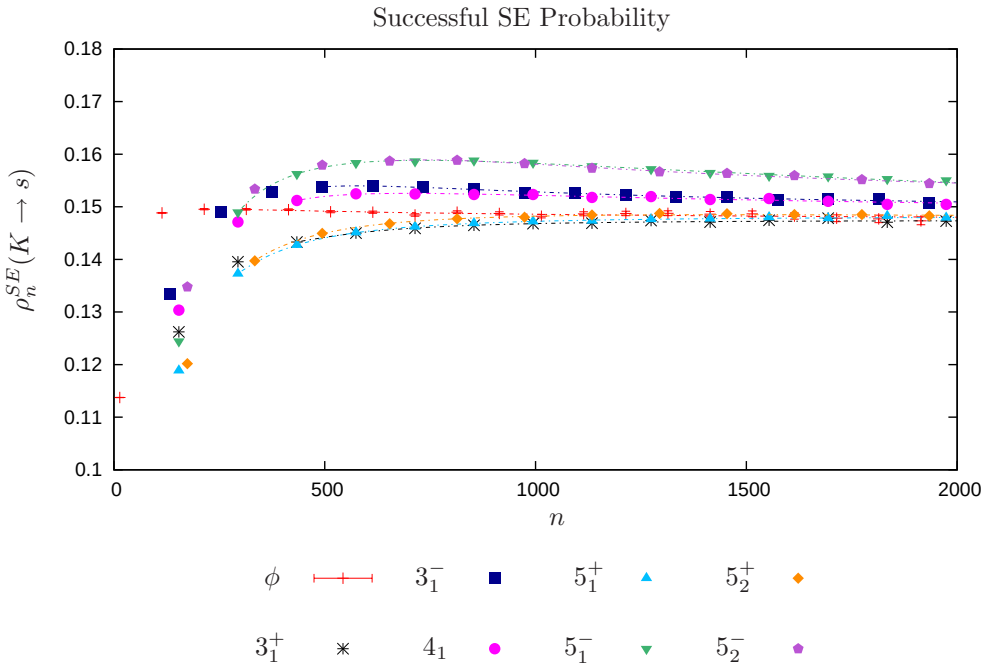


Figure 4.7: A graph of the successful strand exchange probability grouped-n estimates, $\rho_n^{SE}(K \to s)$ for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$ (all blocks start at $n = 14$ and block sizes are given in column five of the tables in B.1). For a strand exchange to occur the Θ-SAP must have the required vertices free, and so must be a Θ_1-SAP. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

4.2 Local Strand Exchange (LSE) Results

LSE results for the simulation specified in Section 3.5 will now be presented. To explore the link transition probabilities associated with the direct-repeat-to-link strand exchange (DLE) and the inverted-repeat-to-knot strand exchange (IKE), first a successful strand exchange (SE) must occur. Therefore, the successful SE probability (Equation 2.14) will be studied and estimates will be presented first, followed by the DLE and IKE results in the following two subsections.

Figure 4.7 shows the grouped-n estimates of the successful SE probability, $\rho_n^{SE}(K \to s)$, for each starting knot type $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. It can be seen from the figure that the probability estimates appear to be going to a constant, $C_{K,s}^{SE} = C_{\phi,s}^{SE}$, as $n \to \infty$, as is expected from the conjectured Equation 2.12. To explore this further, fits are performed using the fitting forms for *Case 2* as seen in Equations 3.34-3.36.

The estimates for the limiting $C_{K,s}^{SE}$ value for $\rho_n^{SE}(K \to s)$, as $n \to \infty$, can be seen in Table 4.3. All the estimates are quite close to each other, *i.e.* within the error bars, with $K = 5_1^-$ slightly lower. To get a better view of the limiting probability Figure 4.8 shows $\rho_n^{SE}(K \to s)$ plotted against $\frac{1}{n}$. This plot is chosen because, when plotted versus $\frac{1}{n}$, the y-intercept becomes the limiting value as $n \to \infty$, allowing the limiting estimate

K	$C_{K,s}^{SE}$ value (S.E)	G.O.F. Λ (df)	p value
ϕ	$0.1460(0.0022)^{(5)}$	38.20(16)	0.0014
3_1^+	$0.1474(0.00022)^{(5)}$	18.66(10)	0.0674
3_1^-	$0.1487(0.0012)^{(5)}$	9.25(11)	0.5992
4_1	$0.1472(0.0025)^{(4)}$	13.73(10)	0.1855
5_1^+	$0.1482(0.000083)^{(5)}$	6.14(9)	0.9087
5_1^-	$0.1347(0.0060)^{(6)}$	12.35(12)	0.4178
5_2^+	$0.1429(0.0040)^{(5)}$	14.49(10)	0.1517
5_2^-	$0.1496(0.0020)^{(5)}$	10.56(8)	0.2281

Table 4.3: The results for the grouped-n estimates of the successful SE probability, $\rho_n^{SE}(K \to s)$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the limiting value as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.34-3.36.

for $C_{K,s}^{SE}$ to be easily seen. In Figure 4.8 it can be seen that the limiting probability for $K = 5_1^-$ is very close to that of the other knot types; therefore, all the knot types show a similar trend and look to be going to the same estimate for the successful SE probability. In the limit of large polygons, a SE is estimated to be successful in $\approx 14.5\%$ of all Θ-SAPs for the investigated knot types.

Recall that in order to perform a DLE or an IKE, a successful strand exchange must occur. Since we now know the estimated successful strand exchange probability, in the next subsection we will study the estimated link transition probabilities for a DLE, followed by the estimated link transition probabilities for an IKE.

4.2.1 DLE Results

Given that a SE can occur (*i.e.* given a Θ_1-SAP), we are now interested in the result that a DLE creates. We first look at *Case 1*: when Θ interacts with the original knot and the original knot changes, and then at *Case 2*: when Θ does not interact with the original knot and the original knot remains intact.

DLE Results for *Case 1*

A pictorial example of this case for $K = 3_1^+$ can be seen in Figure 2.11 (a). In this case, Θ interacts with the original knot. Tables B.1-B.8 in Appendix B show all observed link types after a DLE. For some knot types, several different link types were observed after a DLE, so in order to best present the results, multiple link types are grouped together. Figure 4.9 shows the grouped-n estimates of the probability that the link type after a DLE will be a link which does not contain the original knot for each starting knot type K. This set of links is called Γ_{DLE,C_1}. This set will depend on K, but for simplicity this is ignored. As expected, as the polygon length n increases, the probability that Θ interacts with the original knot decreases, meaning that,

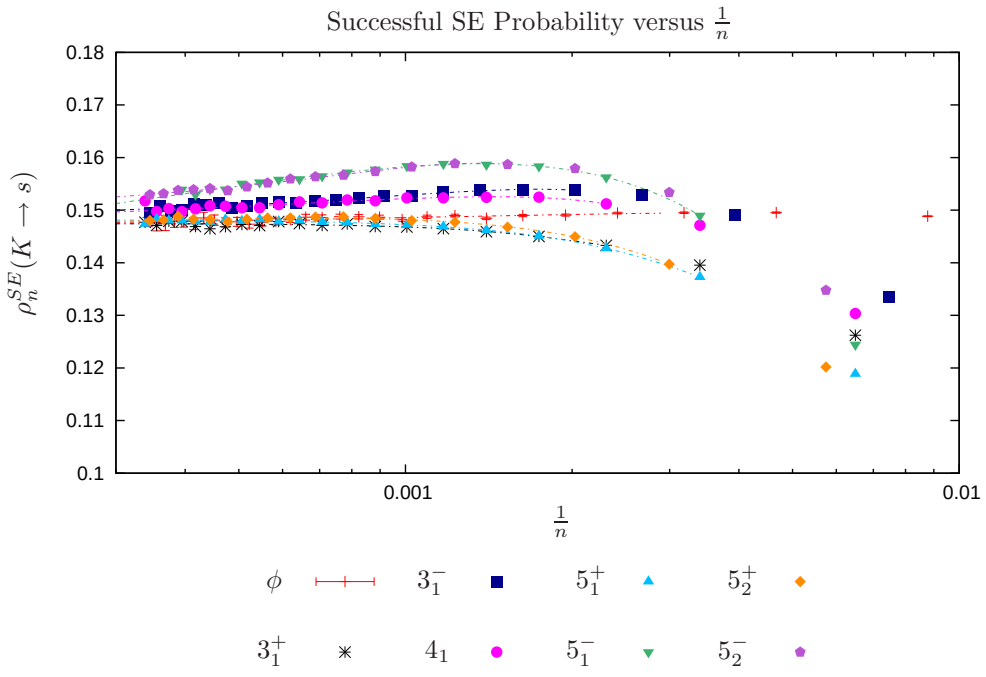


Figure 4.8: A graph of the successful strand exchange probability estimates, $\rho_n^{SE}(K \to s)$ versus $\frac{1}{n}$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

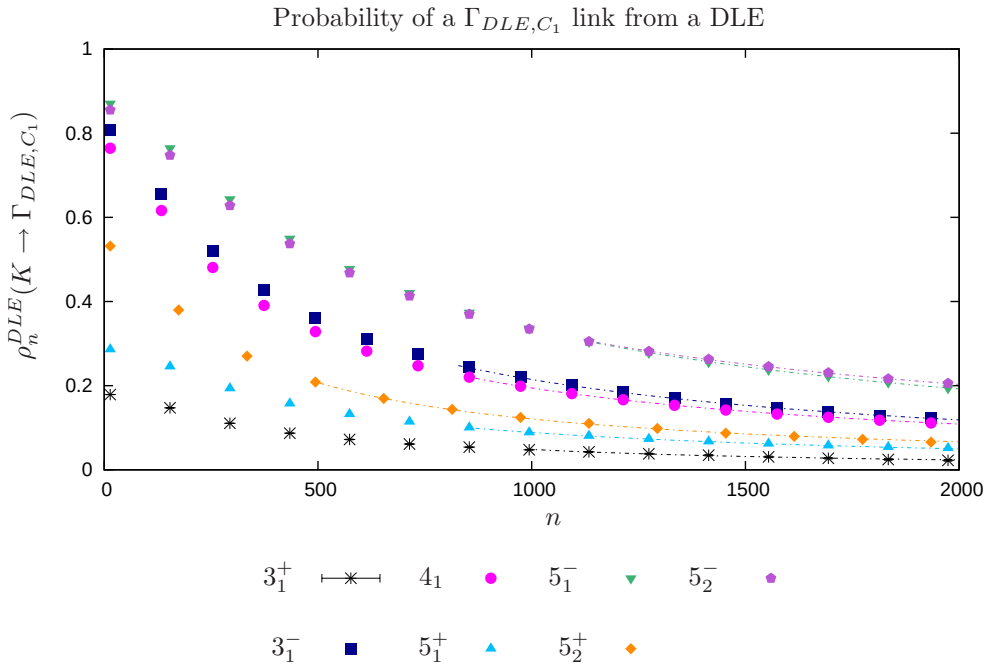


Figure 4.9: A graph of the probability estimates of a Γ_{DLE,C_1} link resulting from a DLE, $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

as polygon length increases, it becomes less likely that the original knot is changed after a DLE.

To further investigate whether the probability estimates of $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ are going to zero as $n \to \infty$, Figure 4.10 shows the grouped-n estimates of $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ plotted with the y-axis on a logarithmic scale. By taking the logarithm of the chosen fitting form, for example $\rho_n = mn^{-t^{SE}_{K,\Gamma_{DLE,C_1}}}$, the graphed estimates become a linear plot. For example $\log(\rho_n) = \log(m) - (t^{SE}_{K,\Gamma_{DLE,C_1}})\log(n)$, so the slope of the graphed equation is $t^{SE}_{K,\Gamma_{DLE,C_1}}$. Therefore, if the logarithm of the fitting forms are parallel, then each is decreasing to zero like $n^{-t^{SE}_{K,\Gamma_{DLE,C_1}}}$ and the exponents, $t^{SE}_{K,\Gamma_{DLE,C_1}}$, must be equal. From Figure 4.10 it can be seen that each knot type has a similar behaviour from the parallel nature of the plots. Table 4.4 shows the limiting estimate for $t^{DLE}_{K,\Gamma_{DLE,C_1}}$ for the fits shown in Figure 4.9 and Figure 4.10.

Most of the estimated values of $t^{DLE}_{K,\Gamma_{DLE,C_1}}$ are quite close to the expected value of 1; however, $K = 5_2^-$ has a lower estimate and $K = 5_2^+$ has a p value < 0.05. Out of the knots types studied, $K = 5_2$ has the largest minimum number of edges, so it is possible that by cutting off n at $N_{max}(K)$, the limiting value is not being seen yet. Indeed, if the cut off point is pushed up by 300 edges, the $K = 5_2^-$ estimate gets much closer to a value of 1 and the $K = 5_2^+$ p value becomes much larger, as seen in the last two rows of Table 4.4.

To gain further evidence that each $t^{DLE}_{K,\Gamma_{DLE,C_1}}$ is equal to a value of 1, a plot of $\log(n\rho_n^{DLE}(K \to \Gamma_{DLE,C_1}))$

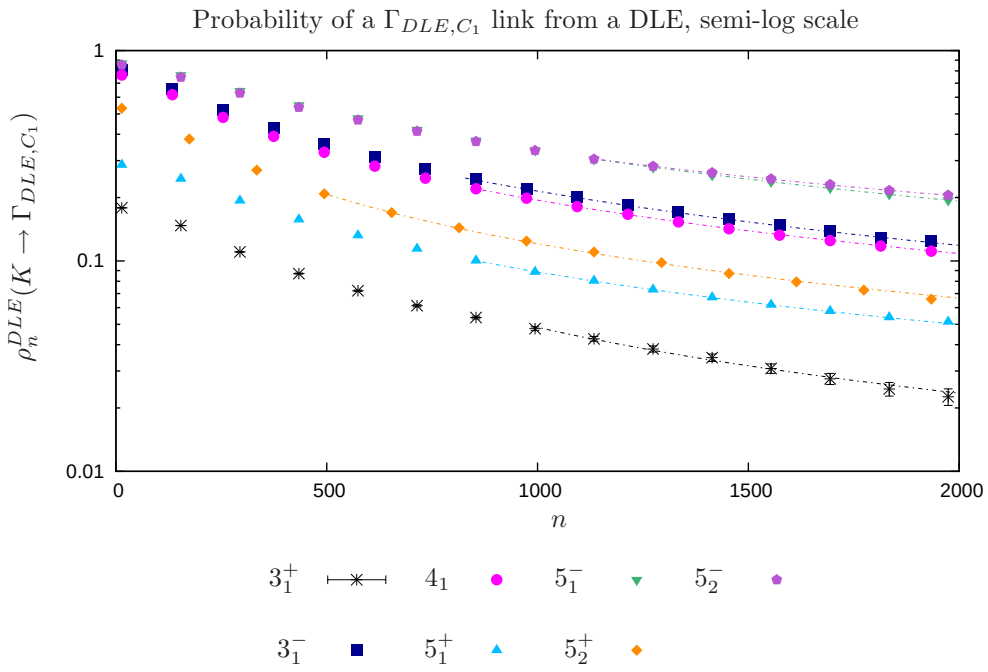


Figure 4.10: A graph of the probability estimates of a Γ_{DLE,C_1} link resulting from a DLE, $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$, on a semi-log scale. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

K	$t^{DLE}_{K,\Gamma_{DLE,C_1}}$ value (S.E)	G.O.F. Λ (df)	p value
3_1^+	$1.0389(0.034)^{(1)}$	13.56(9)	0.1389
3_1^-	$1.0099(0.019)^{(3)}$	0.67(13)	0.9986
4_1	$0.935(0.039)^{(2)}$	1.75(9)	0.9948
5_1^+	$0.9454(0.038)^{(3)}$	2.17(9)	0.9884
5_1^-	$0.9962(0.038)^{(2)}$	0.95(7)	0.9956
5_2^+	$0.9317(0.013)^{(3)}$	28.98(6)	0.0023
5_2^-	$0.8351(0.0048)^{(3)}$	1.22(11)	0.9999
5_2^+ extended	$1.023(0.022)^{(1)}$	5.02(9)	0.8322
5_2^- extended	$0.975(0.017)^{(3)}$	0.29(9)	0.9999

Table 4.4: The results of a Γ_{DLE,C_1} link from a DLE, $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$, for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value $t^{DLE}_{K,\Gamma_{DLE,C_1}}$ as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.30-3.32.

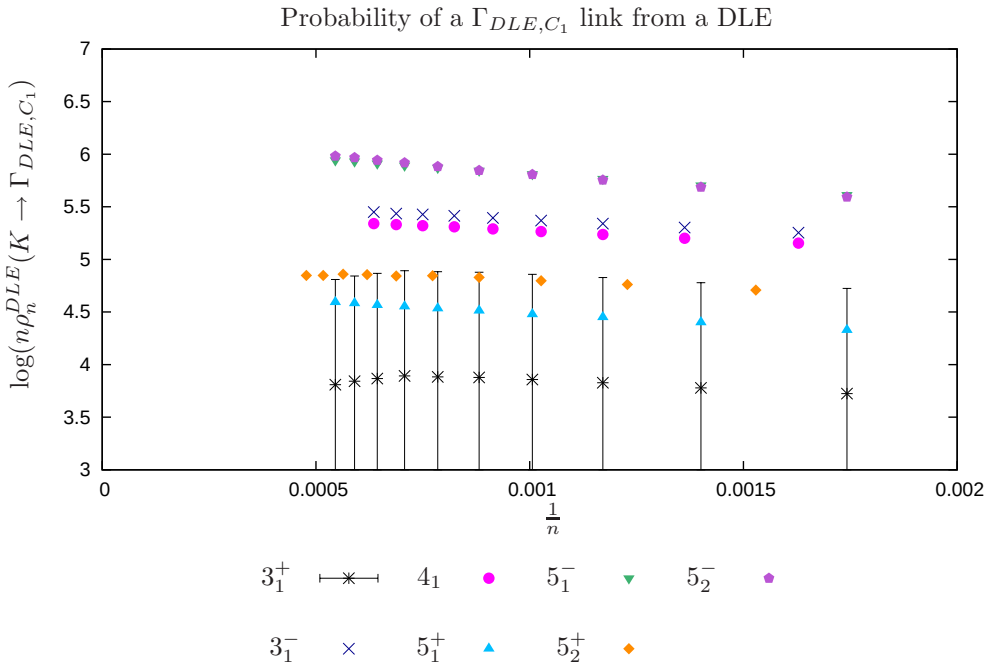


Figure 4.11: A graph of the logarithm of n times the probability estimates of a Γ_{DLE,C_1} link resulting from a DLE versus $\frac{1}{n}$ for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

versus $\frac{1}{n}$ is presented in Figure 4.11. This plot is chosen because a plot of the logarithm of n times the fitting form will result in a linear plot where the y-intercept corresponds to a constant from the fitting form. For example, again for the form $\rho_n = mn^{-t^{SE}_{K,\Gamma_{DLE,C_1}}}$, the plotted equation would be $\log(n\rho_n) = \log(n(mn^{-t^{SE}_{K,\Gamma_{DLE,C_1}}})) = \log(m) + (1 - t^{SE}_{K,\Gamma_{DLE,C_1}})\log(n)$. If $t^{SE}_{K,\Gamma_{DLE,C_1}}$ was the expected value of 1, then the plotted equation would be of the form $\log(m)$, a constant. Consistent with the previous evidence, these estimates appear nearly horizontal with $\frac{1}{n}$ and approaching a constant as $n \to \infty$. This provides strong numerical evidence supporting the conjecture in Equation 2.18.

It should be noted that Figures 4.9, 4.10, and ultimately 4.11 provide a ranking for the knot types K. That is, for each n, the estimates for $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ are largest for $K = 5_1^-, 5_2^-$ and then decrease sequentially with $K = 3_1^-, 4_1, 5_2^+, 5_2^-$, and 3_1^+. Recall that K^- corresponds to the (0) scenario with DLE and K^+ corresponds to the (+1) scenario with DLE (*cf.* Section 2.3), meaning that the class I polygons (K^-) give larger estimates. In order to get all results for K^+, both K^+ and K^- need to be combined. It is expected that after combining then the ranking would go according to crossing number from highest to lowest, *i.e.* 5_1 and 5_2 would be highest, followed by 4_1 and then 3_1.

DLE Results for *Case 2*

A pictorial example of this case with $K = 3_1^+$ can be seen in Figure 2.12 (a). In this case, a two component link is formed at Θ, where the original knot type K is in the decomposition of one of the two components of the link. We know this case becomes dominates as $n \to \infty$, since *Case 1* goes to zero as $n \to \infty$. However, as the analysis in Section 3.5.5 showed, for the lengths that are considered reliable, $\rho_n^{DLE}(K \to \Gamma_{DLE,C_1})$ is greater than zero. This means that *Case 1* will bias the *Case 2* analysis unless dealt with. To eliminate this bias, *Case 1* links are removed from the set of successful strand passage polygons analyzed, giving Equation 3.33 to obtain the estimates.

For *Case 2*, since for large polygons, knots are usually localized within a greater unknot [20, 55], with Θ not interacting with the original knot, it is most likely that a DLE will just close off a section of the large unknotted part, forming a second unknotted component. Though it is possible that a more complicated knot will be formed, the most likely result of a DLE will be that the original knot remains unchanged and a secondary unknot will be created at Θ. The result will then be a link whose two components are the original knot and an unknot created at Θ, either linked or unlinked. As stated earlier, if the two components are not linked together, *i.e.* unlinked, the link will be represented by $L = u(K \cup \phi)$. Figure 4.12 shows the grouped-n estimates of the probability of $L = u(K \cup \phi)$ after a DLE, $\rho_n^{DLE*}(K \to u(K \cup \phi))$. From the figure it can be seen that all the plots appear to be going to the same constant as $n \to \infty$. The estimates for the limiting value $C_{K,u(K\cup\phi)}^{DLE}$ can be seen in Table 4.5. Even though the p value < 0.05 for ϕ, it can be seen that all the estimates for $C_{K,u(K\cup\phi)}^{DLE}$ are ≈ 0.976 and are within error bars of each other. To show further evidence that all knot types are indeed going to the same constant, Figure 4.13 shows plots of the $\rho_n^{DLE*}(K \to u(K \cup \phi))$ estimates versus $\frac{1}{n}$. From this graph it can be seen that all plots are going to the same limiting value within error. This provides strong evidence that $\rho_n^{DLE*}(K \to u(K \cup \phi))$ is going to a constant, $C_{K,u(K\cup\phi)}^{DLE} \approx 0.976$ as $n \to \infty$, confirming the conjecture in Equation 2.20.

Another possibility with the creation of an unknot outside the original knot is that the two components will be linked together, *i.e.* $L = l(K \cup \phi)$. Figure 4.14 shows the grouped-n estimates of the probability of $L = l(K \cup \phi)$ after a DLE, $\rho_n^{DLE*}(K \to l(K \cup \phi))$. From the figure it can be seen that all the plots appear to be going to the same constant, $C_{K,l(K\cup\phi)}^{DLE}$, as expected. To get a better look at the limiting value, the estimates of $\rho_n^{DLE*}(K \to l(K \cup \phi))$ are plotted versus $\frac{1}{n}$ as shown in Figure 4.15. The estimates indeed are seen to be going to the same limiting value, $C_{K,l(K\cup\phi)}^{DLE} \approx 0.022$, within error. This provides further evidence that the conjecture in Equation 2.20 holds.

Therefore, both *Case 2* scenarios considered illustrate that the link transition probabilities resulting from a *Case 2* DLE are going to constants, independent of the starting knot type K, but dependent on the resulting link type L. It is possible that when DLE occurs, it does not create an unknot at Θ, but a more complicated knot, which again could be linked or unlinked with the original knot. In the generated data, examples of this were observed with starting knots $K = 3_1^+$ and $K = 3_1^-$ (*cf.* Table B.2 and B.3); however too few of these knots were observed to make any estimates or say anything conclusive.

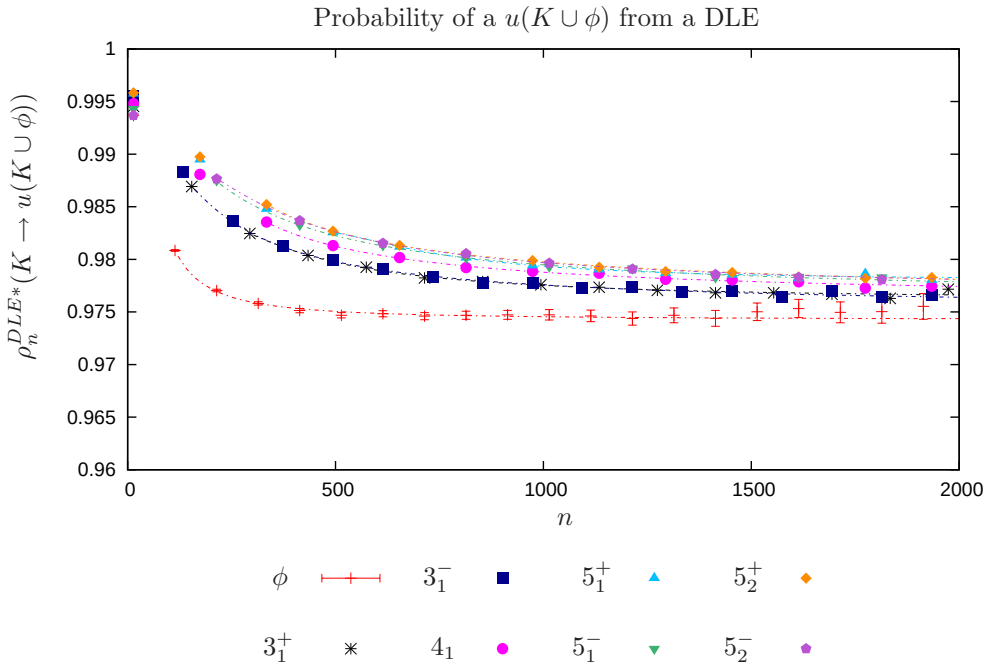


Figure 4.12: A graph of the probability estimates of a $u(K \cup \phi)$ link resulting from a DLE, $\rho_n^{DLE*}(K \to u(K \cup \phi))$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

K	$\boldsymbol{C^{DLE}_{K,u(K\cup\phi)}}$ **value (S.E)**	**G.O.F.** $\boldsymbol{\Lambda}$ *(df)*	***p* value**
ϕ	$0.9743(0.00013)^{(4)}$	26.43(17)	0.048
3_1^+	$0.9761(0.00033)^{(6)}$	6.10(10)	0.807
3_1^-	$0.9753(0.00036)^{(5)}$	4.85(14)	0.938
4_1	$0.9758(0.00032)^{(5)}$	4.19(11)	0.938
5_1^+	$0.9780(0.00015)^{(5)}$	3.51(9)	0.940
5_1^-	$0.9760(0.00058)^{(6)}$	3.50(9)	0.900
5_2^+	$0.9759(0.00031)^{(4)}$	5.14(11)	0.924
5_2^-	$0.9766(0.00046)^{(5)}$	4.54(8)	0.806

Table 4.5: The results of a $u(K \cup \phi)$ link from a DLE, $\rho_n^{DLE*}(K \to u(K \cup \phi))$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.34-3.36.

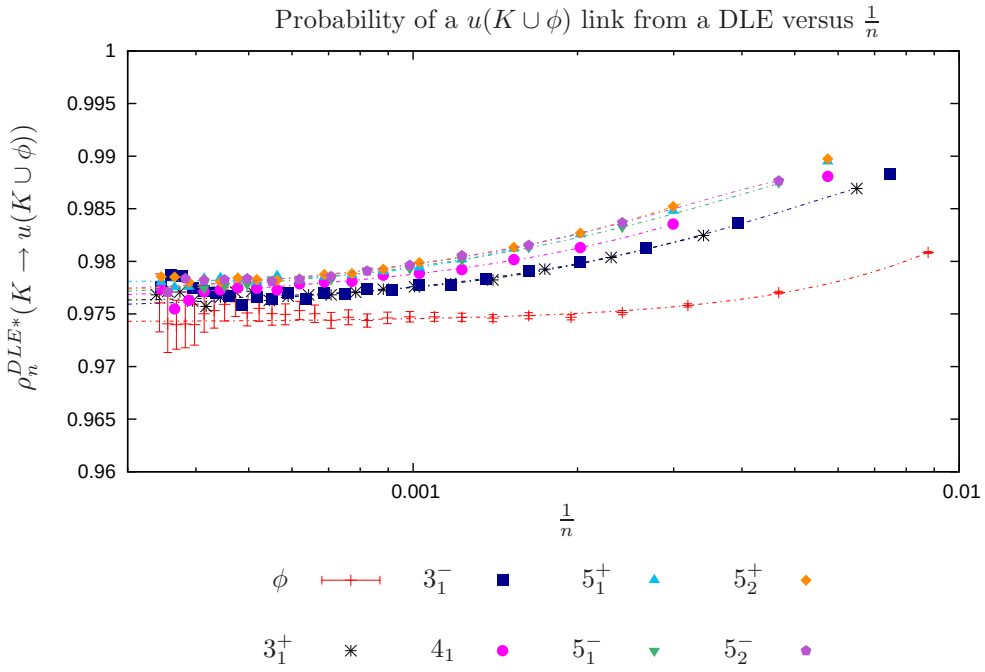


Figure 4.13: A graph of the probability estimates of a $u(K \cup \phi)$ link resulting from a DLE, $\rho_n^{DLE*}(K \to u(K \cup \phi))$ versus $\frac{1}{n}$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

K	$C^{DLE}_{K,l(K\cup\phi)}$ value (S.E)	G.O.F. Λ (df)	p value
ϕ	$0.02386(0.000082)^{(4)}$	12.326(11)	0.3397
3_1^+	$0.02239(0.00015)^{(5)}$	2.17(8)	0.9756
3_1^-	$0.02250(0.00012)^{(5)}$	1.35(12)	0.9981
4_1	$0.02290(0.00028)^{(4)}$	5.45(11)	0.9073
5_1^+	$0.02104(0.00013)^{(4)}$	3.07(9)	0.9299
5_1^-	$0.02306(0.000058)^{(6)}$	3.82(9)	0.8731
5_2^+	$0.02127(0.00017)^{(6)}$	1.41(6)	0.9873
5_2^-	$0.02157(0.00023)^{(5)}$	1.38(6)	0.9671

Table 4.6: The results of a $l(K \cup \phi)$ link from a DLE, $\rho_n^{DLE*}(K \to l(K \cup \phi))$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.34-3.36.

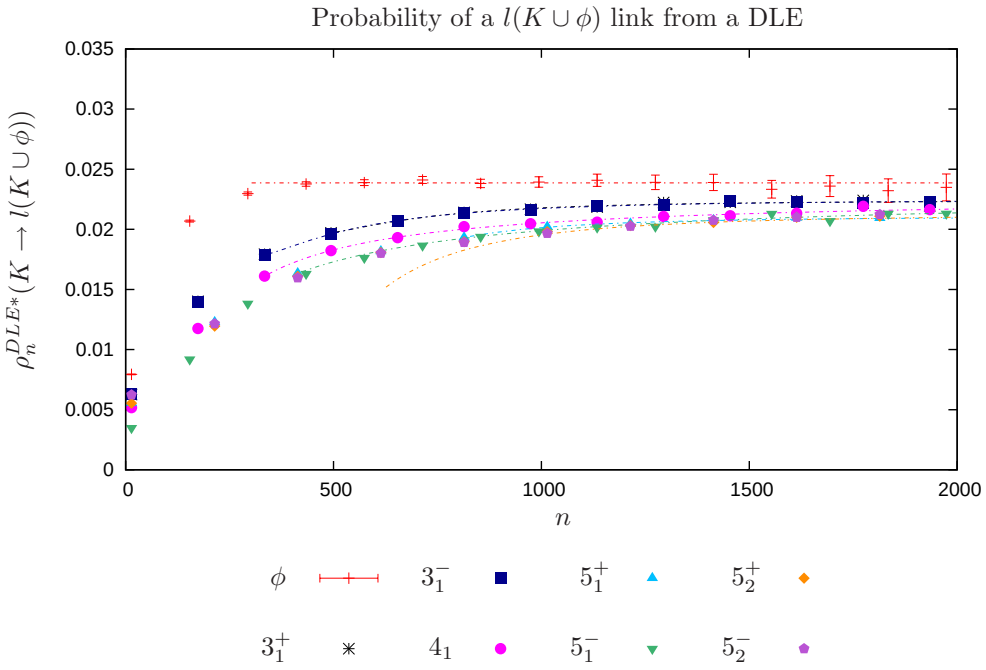


Figure 4.14: A graph of the probability estimates of a $l(K \cup \phi)$ link resulting from a DLE, $\rho_n^{DLE*}(K \to l(K \cup \phi))$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

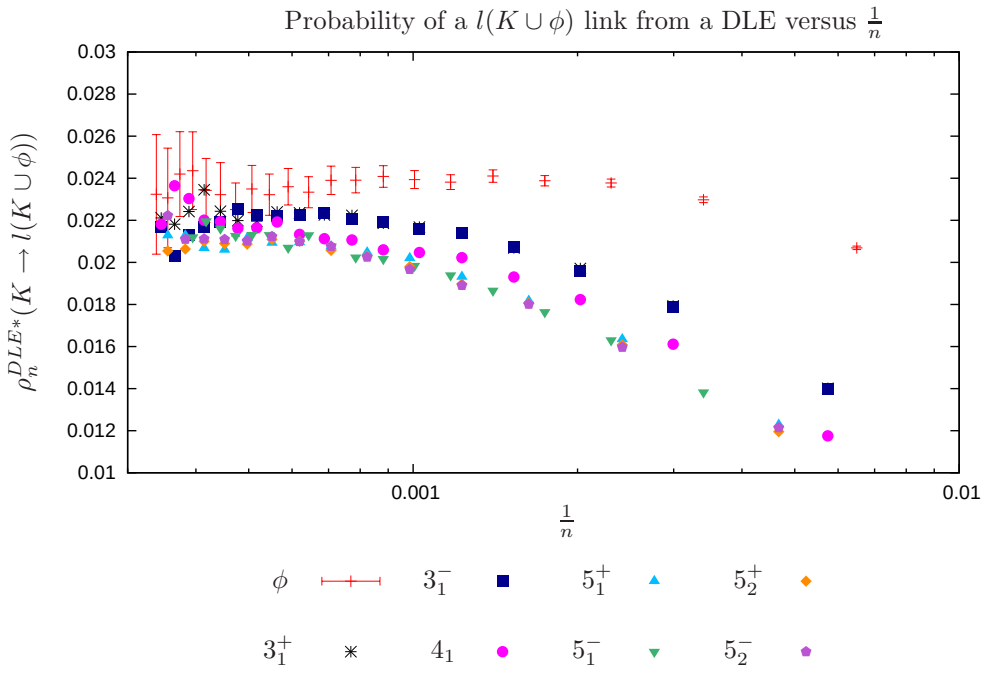


Figure 4.15: A graph of the probability estimates of a $l(K \cup \phi)$ link resulting from a DLE, $\rho_n^{DLE*}(K \to l(K \cup \phi))$ versus $\frac{1}{n}$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

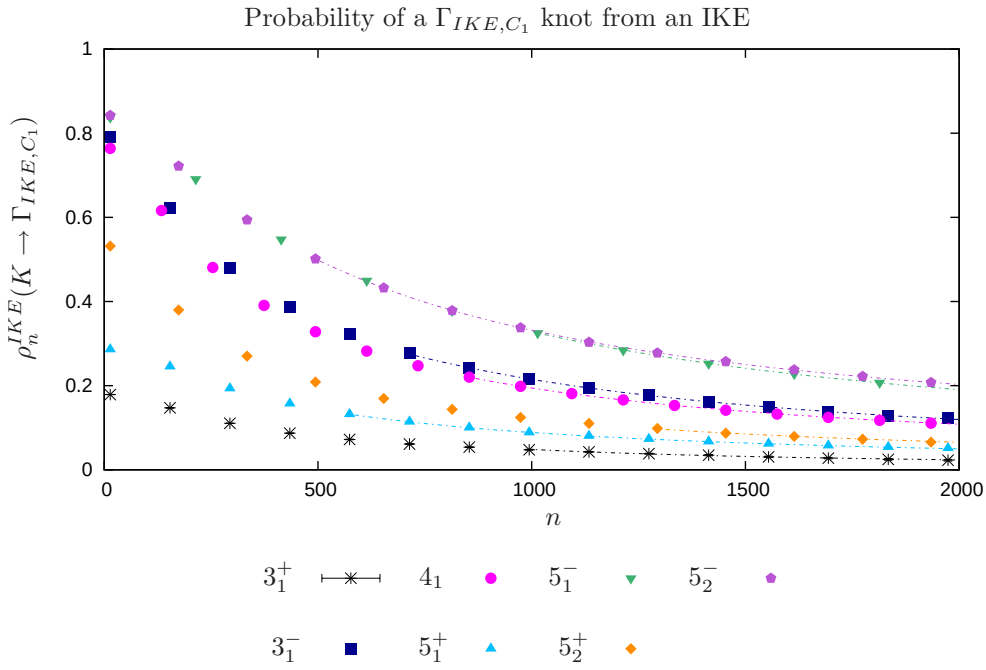


Figure 4.16: A graph of the probability estimates of a Γ_{IKE,C_1} knot resulting from an IKE, $\rho_n^{IKE}(K \to \Gamma_{IKE,C_1})$, for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

4.2.2 IKE Results

Now that DLE results have been looked at, we will turn to IKE. Again we'll look at *Case 1* first, followed by *Case 2*.

IKE Results for *Case 1*

A pictorial example of this case can be seen in Figure 2.11 (b). In this case Θ interacts with the original knot, therefore changing the original knot type. As with DLE, the knots created from an IKE which do not contain the original knot type will be grouped together in the set Γ_{IKE,C_1}. Figure 4.16 shows the grouped-n estimates for the probability that after an IKE the knot will be Γ_{IKE,C_1}, $\rho_n^{IKE}(K \to \Gamma_{IKE,C_1})$. The figure shows that as n increases, the probability that the original knot will be affected decreases, meaning it becomes less likely that Θ interacts with the original knot. As evidence that the $\rho_n^{IKE}(K \to \Gamma_{IKE,C_1})$ estimates are going to zero as $n \to \infty$, Figure 4.17 shows the semi-log plot of the estimated probability versus polygon length n. On this graph the plots appear parallel, indicating they are all going to zero at the same rate, $n^{-t_{K,L}^{\Gamma_{IKE,C_1}}}$. Table 4.7 gives the estimates for $t_{K,\Gamma_{IKE,C_1}}^{IKE}$.

Like the DLE situation, the IKE estimates for $t_{K,\Gamma_{IKE,C_1}}^{IKE}$ are quite close to the expected value of 1, with

K	$t^{IKE}_{K^{IKE},\Gamma_{IKE,C_1}}$ value (S.E)	G.O.F. Λ (df)	p value
3_1^+	$1.0287(0.032)^{(1)}$	12.80(9)	0.1720
3_1^-	$0.9930(0.020)^{(3)}$	0.67(7)	0.9985
4_1	$0.9396(0.030)^{(3)}$	2.14(9)	0.9890
5_1^+	$0.9455(0.028)^{(3)}$	2.15(9)	0.9889
5_1^-	$0.9601(0.028)^{(2)}$	0.59(5)	0.9884
5_2^+	$0.9802(0.020)^{(3)}$	16.29(10)	0.0915
5_2^-	$0.8396(0.0040)^{(3)}$	1.40(10)	0.9992
5_2^+ extended	$0.9845(0.018)^{(1)}$	7.97(9)	0.6320
5_2^- extended	$0.9248(0.026)^{(2)}$	0.34(7)	0.9999

Table 4.7: The results of a Γ_{IKE,C_1} link from an IKE, $\rho_n^{IKE}(K \to \Gamma_{IKE,C_1})$, for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.30-3.32.

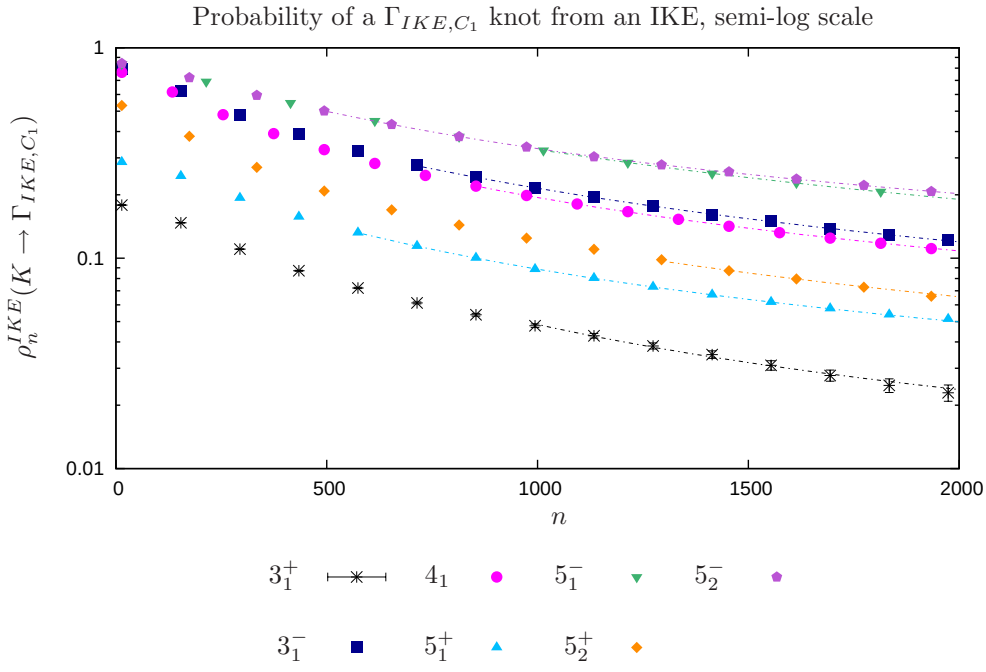


Figure 4.17: A graph of the probability estimates of a Γ_{IKE,C_1} knot resulting from an IKE, $\rho_n^{IKE}(K \to \Gamma_{IKE,C_1})$, for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$, on a semi-log scale. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

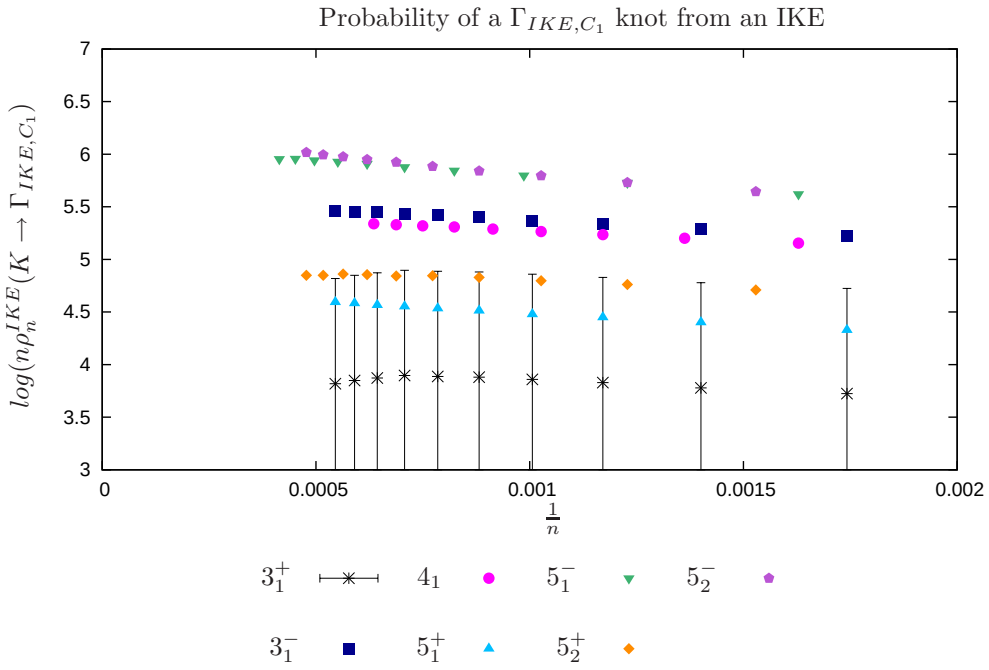


Figure 4.18: A graph of the logarithm of n times the probability estimates of a Γ_{IKE,C_1} link resulting from an IKE versus $\frac{1}{n}$, for $K \in \{3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = 3_1^+$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

$K = 5_2^-$ the lowest, and $K = 5_2^+$ having a p value < 0.05. Again, when the cut off point for reliable data is increased by 300 edges, the $K = 5_2^-$ estimate gets much closer to the value 1 and the $K = 5_2^+$ p value increases, *cf.* Table 4.7. By the same reasoning as for the DLE case, $\log(n\rho_n^{IKE}(K \to \Gamma_{IKE,C_1}))$ is plotted versus $\frac{1}{n}$ in Figure 4.18. This graph shows (almost horizontal) linear plots, supporting that all knot types have the same exponent $t^{IKE}_{K,\Gamma_{IKE,C_1}}$, *i.e.* the exponent is independent of the initial knot type and is equal to one. These results provide strong numerical evidence for the conjecture of Equation 2.19.

Figure 4.18 shows that the ranking from highest probability to lowest of knot type K is the same as that for the DLE case (*cf.* Figure 4.9): $5_1^-, 5_2^-, 3_1^-, 4_1, 5_1^+, 5_2^+$ and 3_1^+. Likewise, if both K^+ and K^- knots were combined it is expected that the rankings would then correspond to the crossing number of the knots, with higher crossing numbers having higher probabilities.

IKE Results for *Case 2*

Recall *Case 2* is when Θ does not interact with the original knot; an example of this can be seen in Figure 2.11 (b). In this case, Θ will typically be in an unknotted section of the polygon. After an IKE the are two possibilities: Θ is still in an unknotted section, and so the whole knot type is unchanged, or a knot can be

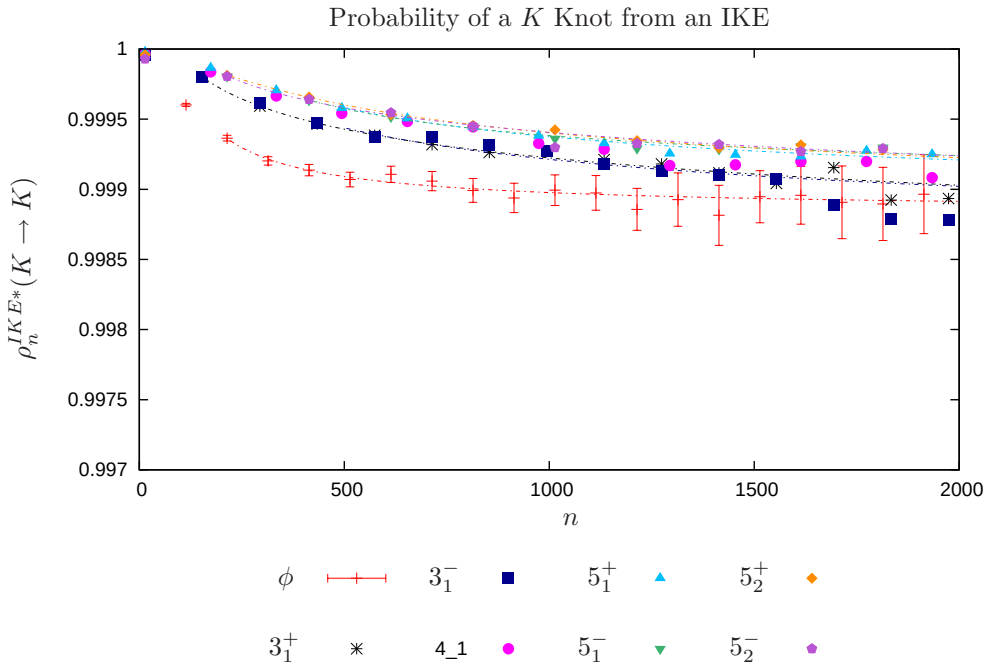


Figure 4.19: A graph of the probability estimates of a K knot resulting from an IKE, $\rho_n^{IKE*}(K \to K)$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

created at Θ, resulting in a composite knot. Figure 4.19 shows a graph of the grouped-n estimates of the probability associated with the first option *i.e.* the after IKE knot type is the same knot type as the original knot, $\rho_n^{IKE*}(K \to K)$ (again once *Case 1*'s bias has been removed). From the graph it can be seen that this is the most likely option to occur. Table 4.8 shows the estimates of $C_{K,K}^{IKE}$ are going to the same constant ≈ 0.998, as $n \to \infty$. Using the statistical package R, fits for $K = 3_1^-, 4_1, 5_2^-$ were not able to be found, likely due to the behaviour of the estimates near the reliability cut off point. Due to this, $N_{max}(K)$ was varied for these knot types to try to achieve fits, but better fits were not found. The computing environment MATLAB was then used and (very poor) fits were found for $K = 3_1^-, 5_2^-$. It is possible that other statistical packages could be used to obtain better fits in the future (*cf.* Section 6.2). In order get a better look at the limiting probabilities to ensure $K = 4_1$, 3_1^- and 5_2^- follow the expected trend, Figure 4.20 shows the grouped-n estimates of $\rho_n^{IKE*}(K \to K)$ plotted versus $\frac{1}{n}$. This graph shows that all the estimates are indeed going to the same limiting value within error (note that the error bars appear so large due to the scale of the graph). This provides strong numerical evidence that the conjecture of Equation 2.21 holds.

As stated, if an IKE does not create an unknot at Θ, then it must create a new knot, making the whole polygon a composite knot. The most likely knot to be created is the simplest non-trivial knot, the trefoil.

K	$C_{K,K}^{IKE}$ value (S.E)	G.O.F. Λ (df)	p value
ϕ	$0.9988(0.000064)^{(4)}$	10.36(16)	0.8473
3_1^+	$0.9951(0.0032)^{(4)}$	7.60(13)	0.8689
3_1^-	$0.9984(0.00097)^{(5)}$	$>> 10^5(12)$	≈ 0
4_1	–	–	–
5_1^+	$0.9988(0.00020)^{(6)}$	7.12(11)	0.7891
5_1^-	$0.9988(0.00032)^{(4)}$	3.36(8)	0.9099
5_2^+	$0.9986(0.00049)^{(6)}$	4.79(8)	0.7793
5_2^-	$0.9957(0.0085)^{(4)}$	34.44(9)	0.00016

Table 4.8: The results of a K knot from an IKE, $\rho_n^{IKE*}(K \to K)$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value as $n \to \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.34-3.36.

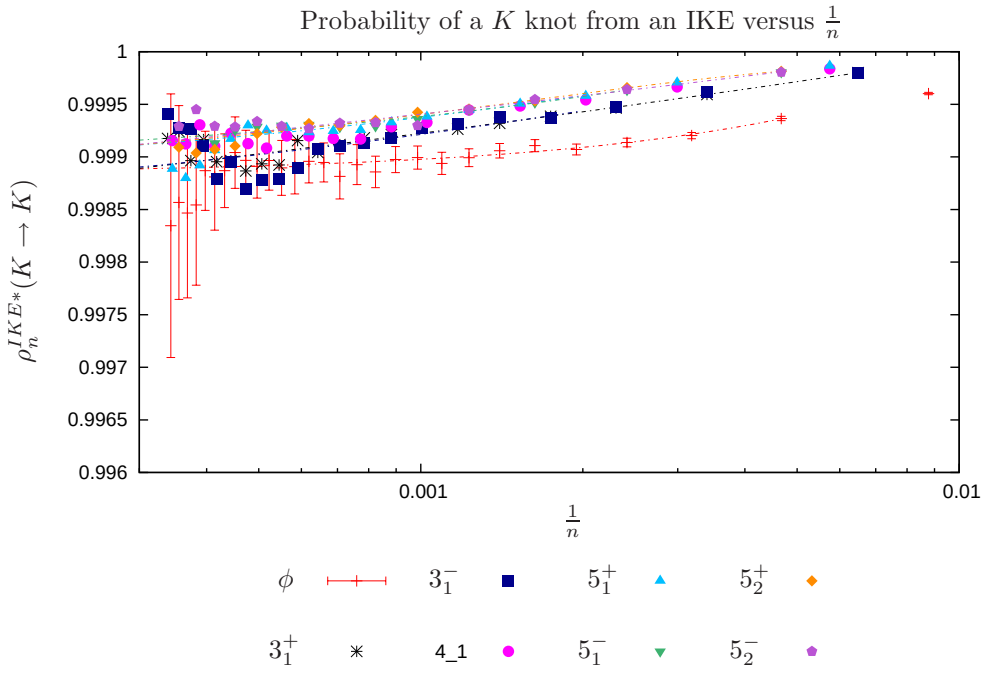


Figure 4.20: A graph of the probability estimates of a K knot resulting from an IKE, $\rho_n^{IKE*}(K \to K)$ versus $\frac{1}{n}$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

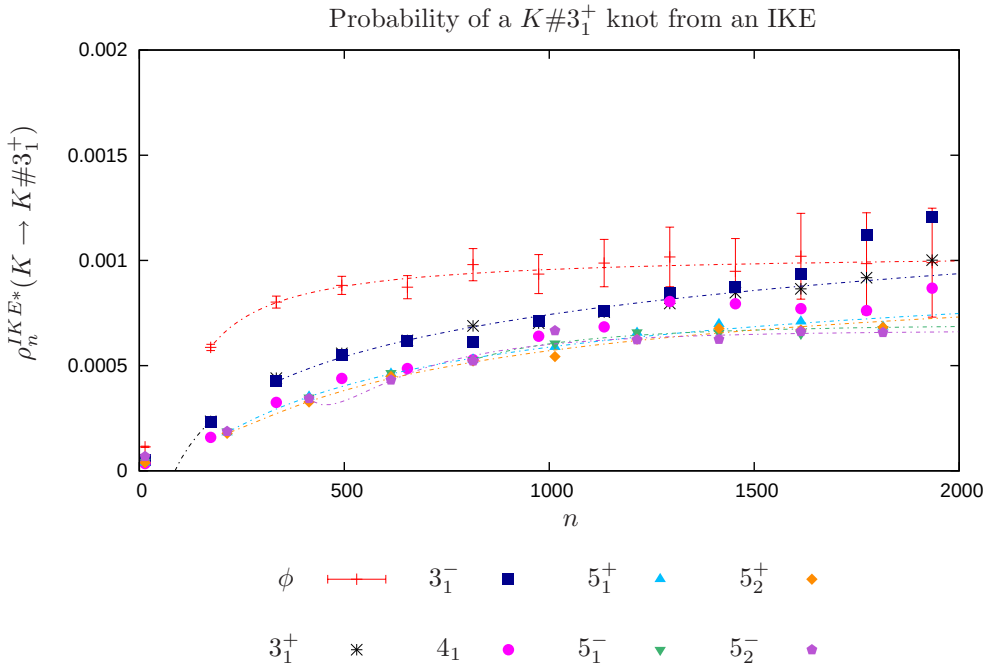


Figure 4.21: A graph of the probability estimates of a $K\#3_1^+$ knot resulting from an IKE, $\rho_n^{IKE*}(K \to l(K\#3_1^+))$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

Figure 4.21 shows the grouped-n estimates of the probability that the resulting polygon knot is $K\#3_1^+$, $\rho_n^{IKE*}(K \to K\#3_1^+)$. Once again all the estimates appear to be going to a constant. To better see this, Figure 4.22 shows the estimate of $\rho_n^{IKE*}(K \to K\#3_1^+)$ plotted versus $\frac{1}{n}$. Table 4.9 shows that the estimates of $C^{IKE}_{K,K\#3_1^+}$ are going to the same constant ≈ 0.001 within error, as $n \to \infty$. Note that 3_1^- has a p value < 0.05, but it appears to follow the trends of the other knots in the figures. Again, this provides strong evidence for the conjecture in Equation 2.21.

It is possible that a knot more complicated than the trefoil can be created at Θ and this was observed for multiple starting knot types; however, too few were observed to say anything conclusive.

K	$C^{IKE}_{K,K\#3_1^+}$ value (S.E)	G.O.F. Λ (df)	p value
ϕ	$0.001040(0.000043)^{(4)}$	4.42(13)	0.8816
3_1^+	$0.002842(0.0011)^{(4)}$	4.70(11)	0.9448
3_1^-	$0.00784(0.067)^{(4)}$	$>> 10^5(9)$	≈ 0
4_1	$0.0008278(0.000034)^{(5)}$	3.13(7)	0.9256
5_1^+	$0.001394(0.00063)^{(5)}$	3.46(8)	0.9024
5_1^-	$0.0006992(0.000026)^{(4)}$	0.92(6)	0.9884
5_2^+	$0.001219(0.00041)^{(5)}$	4.04(6)	0.8531
5_2^-	$0.0006686(0.000034)^{(4)}$	6.04(7)	0.5455

Table 4.9: The results of a $K\#3_1^+$ link from an IKE, $\rho_n^{IKE*}(K \rightarrow K\#3_1^+)$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. The first column is the starting knot type K. The second column is the estimated limiting value as $n \rightarrow \infty$ with the associated standard error. The third column is the goodness of fit test statistic, Λ, with the associated degrees of freedom (df), and the corresponding p value in the fourth column. Superscripts on the estimates represent the fitting form corresponding to Equations 3.34-3.36.

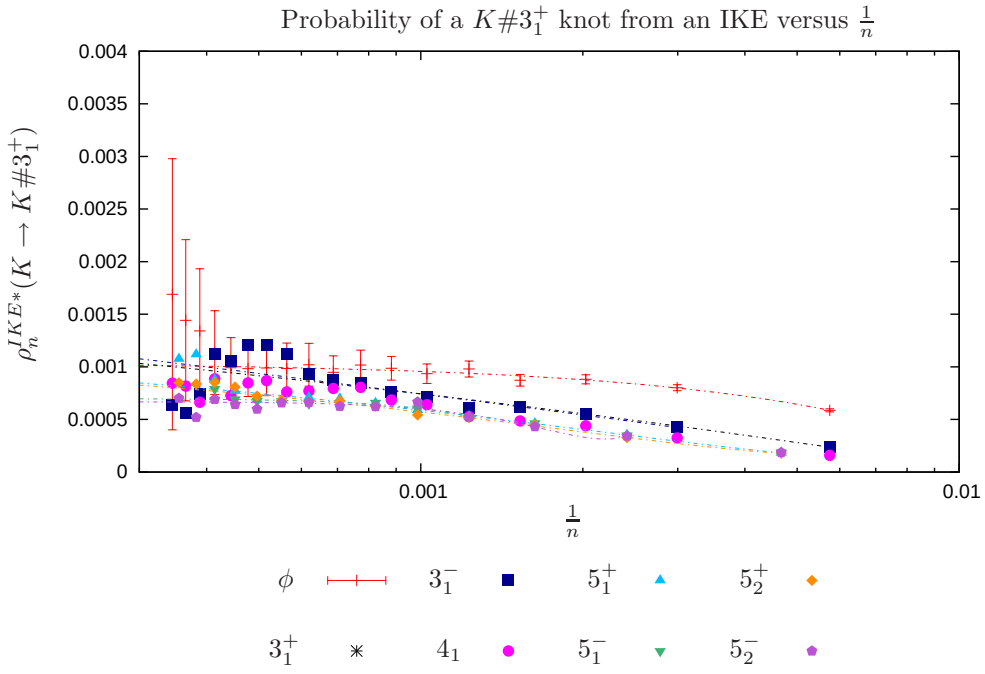


Figure 4.22: A graph of the probability estimates of a $K\#3_1^+$ knot resulting from an IKE, $\rho_n^{IKE*}(K \rightarrow K\#3_1^+)$ versus $\frac{1}{n}$, for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Error bars representing 95% confidence intervals are shown for $K = \phi$, which is greater than or equal to the widths of all other confidence intervals. Only those 95% confidence intervals whose widths are larger than the symbol representing the point estimate can be seen.

4.2.3 SE Results Summary

All the above analyses support the conjectured asymptotic forms for the link transition probabilities presented in Equations 2.18-2.21:

$$\rho_n^{DLE}(K \to L) \sim \left(\frac{B_{DLE(K,L)} A_\phi}{B_{\Theta_1} A_L} \right) \frac{1}{n^{f_K}}$$

$$\rho_n^{IKE}(K \to K') \sim \left(\frac{B_{IKE(K,K')} A_\phi}{B_{\Theta_1} A_K} \right) \frac{1}{n^{f_K}}$$

$$\rho_n^{DLE}(K \to L) \sim \frac{B_{DLE(\phi,L)}}{B_{\Theta_1}}$$

$$\rho_n^{IKE}(K \to K') \sim \frac{B_{IKE(\phi,K'')}}{B_{\Theta_1}}.$$

When in *Case 1* (the original knot is affected), the link transition probability estimates are shown going to zero for both DLE and IKE, as the length of the Θ-SAP increases, *i.e.*

$$\rho_n^{SE}(K \to L) \to 0 \qquad \text{as } n \to \infty \quad \text{like } n^{-t_{K,L}^{SE}}. \tag{4.1}$$

The analysis supports that for both DLE and IKE, $t_{K,L}^{SE} = 1$, with a summary of the estimates shown in Table 4.10.

When in *Case 2* (the original knot is not affected), the link transition probability estimates go to a constant dependent on the newly created link, as the length of the Θ-SAP increases, *i.e.*

$$\rho_n^{SE*}(K \to L) \to C_{K,L}^{SE} \qquad \text{as } n \to \infty. \tag{4.2}$$

The results show that, for both DLE and IKE $C_{K,L}^{SE} > 0$ and this value is unique for each L, *i.e.* $C_{K,L}^{SE} = C_{\phi,L}^{SE}$. A summary of estimates of $C_{K,L}^{IKE}$ are found in Table 4.10.

The conjectured forms in Equations 2.18-2.21 are based on Cheston *et al.*'s [20] forms in Equations 2.10 and 2.11, which were based on Orlandini *et al.*'s [74] original basic conjectured form for the number of n-edge polygons with fixed knot-type K. Therefore, this work with the new LSE model provides new evidence supporting the form $p_n(K) \sim A_K n^{\alpha_\phi + f_K}$ as $n \to \infty$.

K	$t^{DLE}_{K,\Gamma_{DLE,C_1}}$	$C^{DLE}_{K,u(K\cup\phi)}$	$C^{DLE}_{K,l(K\cup\phi)}$	$t^{IKE}_{K,\Gamma_{IKE,C_1}}$	$C^{IKE}_{K,K\#\phi}$	$C^{IKE}_{K,K\#3_1^+}$
ϕ	_	0.9743	0.02386	_	0.9988	0.001040
		(0.00013)	(0.000082)		(0.000064)	(0.000043)
3_1^+	1.0389	0.9761	0.02239	1.0287	0.9951	0.002842
	(0.034)	(0.00033)	(0.00015)	(0.032)	(0.0032)	(0.0011)
3_1^-	1.0099	0.9753	0.02250	0.9930	0.9984	0.00784
	(0.019)	(0.00036)	(0.00012)	(0.020)	(0.00097)	(0.067)
4_1	0.9354	0.9758	0.02290	0.9396	_	0.0008278
	(0.039)	(0.00032)	(0.00028)	(0.030)		(0.000034)
5_1^+	0.9458	0.9780	0.02104	0.9455	0.9988	0.001394
	(0.038)	(0.00015)	(0.00013)	(0.028)	(0.00020)	(0.00063)
5_1^-	0.9962	0.9760	0.02306	0.9601	0.9988	0.0006992
	(0.038)	(0.00058)	(0.000058)	(0.028)	(0.00032)	(0.000026)
5_2^+	0.9317	0.9759	0.02127	0.9802	0.9986	0.001219
	(0.013)	(0.00031)	(0.00017)	(0.020)	(0.00049)	(0.00041)
5_2^-	0.8351	0.9766	0.02157	0.8396	0.9957	0.0006686
	(0.0048)	(0.00046)	(0.00023)	(0.0040)	(0.0085)	(0.000034)

Table 4.10: Summary of results for link transition probabilities estimates for both DLE and IKE for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Each estimate is followed by its associated standard error.

Additionally, it should be noted that starting with $K = \phi$ the only observed links after a DLE are: $0_1^2, 2_1^2, 4_1^2, 6_1^2, 7_7^2, 7_8^2, l(3_1^+ \cup \phi)$ and $l(3_1^- \cup \phi)$; and the only observed knots after an IKE are: $\phi, 3_1^+, 3_1^-, 5_1^+, 5_1^-, 5_2^-$ and 8_{20}^-. These link types agree with previous results on the number of recombinase steps to reach specific link types [72]. All the knots observed are prime knots which can be unknotted with one crossing change, however, it is notable that no 4-, 6-, and 7-crossing prime knots are observed. It is expected that it is possible to fully characterize both DLE and IKE through the use of tangle calculus. This characterization would allow all possible resulting links to be determined; however this is beyond the scope of this thesis, see Section 6.2 for future possibilities.

4.3 Comparing LSE Results to LSP Results

In this section, results from the new LSE recombinase model will be compared to the previous LSP topoisomerase model.

The LSP model was created to model topoisomerases' action on DNA via a strand passage at the fixed Θ structure, thus creating the possibility for changing the knot type of the initial Θ-SAP (*cf.* Section 2.2).

This is very similar to an IKE action of the LSE model, allowing the two models to be easily compared. The tables in Section C.1 compare some interesting properties between LSP and LSE models.

Tables C.1, C.4, C.7, C.10, C.13, C.16, C.19 and C.22 show the resulting knot types of a SP compared to the resulting knot types of an IKE for each starting knot type $K = \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. In these tables the first column shows the knot type from a strand passage in the LSP model and the second column shows the knot type of an IKE in the LSE model. The third column shows the number of polygons for which this combination was found, with column four showing the percentage out of the total $560,000,000$ sampled Θ-SAPs. Column five shows the smallest Θ-SAP size with the given combination.

Tables C.2, C.5, C.8, C.11, C.14, C.17, C.20 and C.23 show the observed link types from DLE when a SP keeps the knot type the same. Column one shows the SP knot type (the same as the starting knot), and column two shows the observed link types resulting from a DLE. The third column shows the counts of this particular combination, with the percentage of this set shown in column four. Column five displays the smallest knot of this combination. It can be seen that for all starting knot types K, over 99% of the links are the starting knot unlinked with an unknot, *i.e.* $u(K \cup \phi)$, with very small proportions of other links.

Tables C.3, C.6, C.9, C.12, C.15, C.18, C.21, and C.24 show the observed link types of a DLE strand exchange and the corresponding observed knot types of a SP with the LSP model. Column one shows the resulting link type of a DLE, and column two shows the resulting knot type of a SP. Column three shows the count of the number of Θ-SAP polygons with this combination, and column four shows the sample percentage for the combination. Again, column five shows the smallest polygon length of such a combination.

The LSP and LSE both use the Θ structure as their location to act on the Θ-SAP, however the structures they use to perform their move differ. Despite this there are similarities across the models. As noted earlier, in *Case 1* both a DLE and IKE yield the same rankings of the knot types. This ranking for the LSE model matches that of the LSP model. Specifically, Figures 4.11 and 4.18 show that $K = 5_1^-$ and 5_2^- have the highest values, with $3_1^-, 4_1, 5_2^+, 5_1^+, 3_1^+$ following. This ranking matches that of Figure 6 in [20]. This similarity is expected since the ranking is due to the minimum size of the knots, the freedom of the knots and the chirality. For full details see [20].

For *Case 2*, with the LSP model it was found that $\rho_n^{SP}(K \to K\#\phi) > \rho_n^{SP}(K \to K\#3_1) > \rho_n^{SP}(K \to K\#4_1)$ [20]. This inequality also holds for the LSE model, $\rho_n^{IKE}(K \to K\#\phi) > \rho_n^{IKE}(K \to K\#3_1)$, while not enough data was generated to analyze $\rho_n^{IKE}(K \to K\#4_1)$. In Figures 4.19 and 4.21 it can be seen that different knot types, K, with the same number of crossings have estimated knot transition probabilities equal to each other within error, *i.e.* the data points for the three crossing knots and for the five crossing knots are very close together. This is the same trend that was found for the LSP model [20]. Further, it appears that the higher the number of crossings that the starting knot has, the more likely it is to stay the same knot type for a finite n; however, when $n \to \infty$ the probabilities are equal.

The structure used to perform the SP or SE does come into play when looking at the limiting $C_{K,L}$. From [20], for a SP, two structures were used, Θ and a larger symmetric structure S. For Θ, $C_{K,K}^{\Theta} \approx 0.9780$ and

$C^{\Theta}_{K,K\#3_1^+} \approx 0.0208$, while for S, $C^{S}_{K,K} \approx 0.998$ and $C^{S}_{K,K\#3_1^+} \approx 0.0015$ due to the symmetric structure being larger and needing more free vertices. For a DLE $C^{DLE}_{K,u(K\cup\phi)} \approx 0.975$ and $C^{DLE}_{K,l(K\cup\phi)} \approx 0.022$, while for an IKE, $C^{IKE}_{K,K} \approx 0.998$ and $C^{IKE}_{K,K\#3_1^+} \approx 0.001$. Interestingly, the SP Θ values compare quite closely to the SE DLE values, while the SP S values compare closely to the SE IKE values. At this point this appears to be just a curious coincidence.

So, even though the LSP and LSE are different models there exist common trends that both follow, namely the asymptotic behaviour as $n \to \infty$ satisfy conjectures from polymer scaling theory. The asymptotic trends are the same across different structures, however, some limiting values are structure dependent.

4.4 Chapter 4 Summary

In this chapter the analysis techniques outlined in Chapter 3 were performed for the LSE model. For each starting knot type $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$: the maximum n that is considered reliable was determined, $N_{max}(K)$; the time to equilibrium, τ_{exp} was determined; and the appropriate grouped-n block size was determined for link transition probabilities. Following this, the grouped-n estimates for link transition probabilities were presented, first for DLE and then for IKE. For both types of strand exchange, the conjectured forms from Chapter 2 were found to hold. For *Case 1*, the $t^{SE}_{K,L}$ in Equations 3.30-3.32 were found to go to 1 as $n \to \infty$, and for *Case 2*, the $C^{SE}_{K,L}$ in Equations 3.34-3.36 were found to go to nonzero positive constants dependent only on the strand exchange performed and the resulting link type and not on the starting K. This work numerically supports the new conjectured forms in Equations 2.18-2.21 and supports Cheston *et al.*'s [20] conjectured forms (Equations 2.8 and 2.9) and Orlandini *et al.*'s [75] original form (Equation 2.4).

The next chapter will focus on a specialized subset of Θ-SAPs which will be compared to biologically and mathematically motivated results.

Chapter 5

'Equal-Length' Polygon Results

In Section 1.1.3 it was said that *dif* sites, in the termination region of the cell, remain localized at mid-cell throughout the cell replication and segregation process [73]. This localization allows XerCD recombination to act at *dif* sites in order to resolve chromosome dimers to monomers. As seen in Figure 1.14, the *dif* sites split the DNA into two halves which are equal in length. For this reason, it is of interest to look at the subset of Θ-SAPs where Θ is connected with two equal-length self-avoiding walks (SAWs) to form a polygon, in order to compare to the biologically relevant situation. This will be the focus of this chapter.

The chapter will begin by determining what 'equal-length' SAWs will mean in the LSE model and once this is established, the DLE results for these 'equal-length' SAWs will be investigated, followed by the results for the IKE.

5.1 Polygons with 'Equal-Length' Self-Avoiding Walks (SAWs)

Since a subset of the Θ-SAP data is of interest, having enough reliable data may become an issue. To account for this, and to allow some leeway for a knot to form on one side (or the other), Θ-SAPs will arbitrarily be considered to have 'equal-length' SAWs if the two SAWs connected to Θ have lengths within 24-edges of each other (note that the smallest trefoil with Θ is 24-edges). A Θ-SAP will be called an equal-length SAP (or polygon) if the lengths of the two SAWs connected to Θ are exactly equal, and will be called an 'equal-length' SAP (or polygon) if the lengths of the two SAWs connected to Θ are within 24-edges. A Θ-SAP will be called a variable-length SAP (or polygon) if any lengths for the two SAWs connected to Θ are allowed.

For each of the eight starting knot types, Table 5.1 shows the number of equal-length polygons, the number of 'equal-length' polygons and their corresponding proportions of the total sampled polygons. From this table it can be seen that there are many more 'equal-length' polygons than equal-length polygons. Both counts for $K = \phi$ are the highest, due in part to the unknot having no minimum edge requirement beyond the necessary 14-edges to create a Θ-SAP. Even though looking at 'equal-length' polygons greatly increases the sizes of the samples, having enough reliable data is still an issue. For the subset of 'equal-length' polygons, the largest $N_{max}(K)$ value is for knot types $K = 5_1^-$ and $K = 5_2^-$ at about 500 edges, with the corresponding values for other knot types being significantly lower. Due to this limitation on the availability of reliable data, only general trends for this subset will be looked at in order to relate to biological experiments and

mathematical results (*e.g.* tangle analysis).

K	Count of Equal-length SAPs	Percentage	Count of 'Equal-length' SAPs	Percentage
ϕ	18,556,986	3.31	158,489,204	28.302
3_1^+	1,551,975	0.277	58,044,294	10.365
3_1^-	4,704,649	0.840	99,987,989	17.855
4_1	2,977,039	0.532	65,309,971	11.662
5_1^+	1,696,678	0.303	35,217,976	6.289
5_1^-	3,166,876	0.566	76,405,683	13.644
5_2^+	1,444,072	0.258	30,532,269	5.452
5_2^-	3,166,422	0.565	69,680,701	12.443

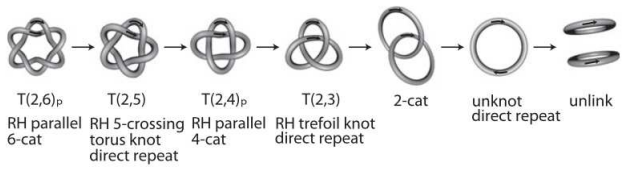
Table 5.1: A comparison of the number of equal-length polygons and the number of 'equal-length' polygons for $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. Percentage columns show the percentage of all sampled polygons which fit into each category.

The remainder of this section will be split into two sections. The first will look at the 'equal-length' polygon results for DLE and comparing these results to the tangle analysis unlinking pathway. The second section will look at 'equal-length' polygon results for IKE to determine if the knot type is simplified or becomes more complex. For both analyses polygons with lengths less than ≈1600 are investigated, meaning that each side is ≈800. The trends seen at these lengths will differ from the trends expected for large lengths (*i.e.* as $n \to \infty$), since the asymptotics from Chapter 4 are expected to also hold for 'equal-length' SAPs; however, these smaller length polygons are expected to be relevant to sizes in DNA experiments.

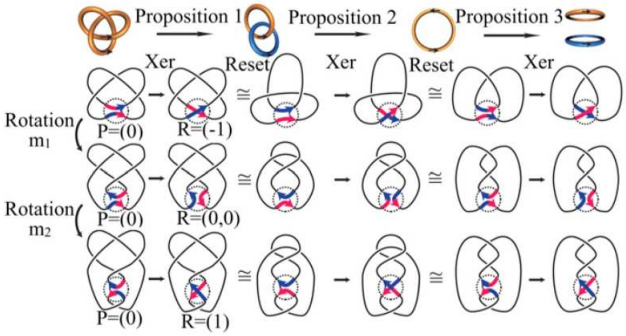
5.2 DLE 'Equal-Length' Polygon Results

Figures 1.28 and 2.10 will be repeated here in Figure 5.1 (a) and (b) for the benefit of the reader as they will be referred to multiple times throughout this section.

Figure 5.1 (b) shows the unlinking path for right-handed (positive) $2m$-catenanes, *i.e.* two monomers which are intertwined together with $2m$ positive crossings. This pathway shows a path of simplification from a link to a knot to a link in order to unlink a $2m$-catenane to the unlink. In Section 2.3 (*cf.* Equations 2.15-2.17) it was explained that link transition probabilities can be obtained from the LSE model, therefore we can investigate recombination with the LSE model and see if this unlinking pathway holds. The LSE model performs only a single recombinase strand exchange at a time, so the whole pathway from link to knot to link cannot be studied; however, single steps which start with knots can be investigated. Five steps will



(a)



(b)

Figure 5.1: (a) The unlinking path for a right-handed (positive) $2m$-catanene. (b) A detailed picture of the last four steps of (a). This shows that there are three possible Xer recombination scenarios when creating a link: $(+1), (-1), (0)$. Both these pictures show that the complexity of the knot is reduced with each step. Reproduced from [95] in accordance with http://www.pnas.org/site/aboutpnas/rightperm.xhtml.

be looked at here: $\phi \rightarrow$ the unlink, $3_1 \rightarrow$ the Hopf link, and each of $4_1, 5_1$ and $5_2 \rightarrow 4_1^2$. As per Section 2.3, looking at one polygon class for the LSE model does not give all the recombination results for a chiral knot K, so for chiral knots the positive knot will be shown, followed by the negative knot, then a combination of the two to get all results for the positive knot. For achiral knots only one graph will be shown.

First, the unknot to the unlink transition ($\phi \rightarrow 0_1^2$), (the last transition of Figure 5.1 (b)) will be investigated. Figure 5.2 shows the probability estimates for the transition $\phi \rightarrow 0_1^2$ after a DLE. The unknot is an achiral knot, meaning performing DLE on $K = \phi$ gives shows results for both polygon classes. It can be seen from Figure 5.2 that the 0_1^2 probability is not 1, this is because the recombinase move that is performed by the LSE is not specifically selected for the topology of the knot as in Figure 5.1 (b). Therefore it is not guaranteed that the move performed will take you solely 'down' the unlinking path (meaning knot or link simplification), it is also possible to go 'up' the pathway (meaning knot complication). It can be seen that when moving from variable-length polygons (solid symbols) to 'equal-length' polygons (open symbols) the estimated probability of transitioning to the unlink, 0_1^2, drops while the estimated probability of the Hopf link (2_1^2), rises; however the unlink remains the dominant result in both cases. Despite the possibility of going 'up' the unlinking pathway, the DLE results are strongly dominated by the link corresponding to moving 'down' the pathway, indicating that this unlinking pathway is the most probable result.

Next, when starting with the trefoil, $K = 3_1$, the unlinking path says that the Hopf link (2_1^2) should be the

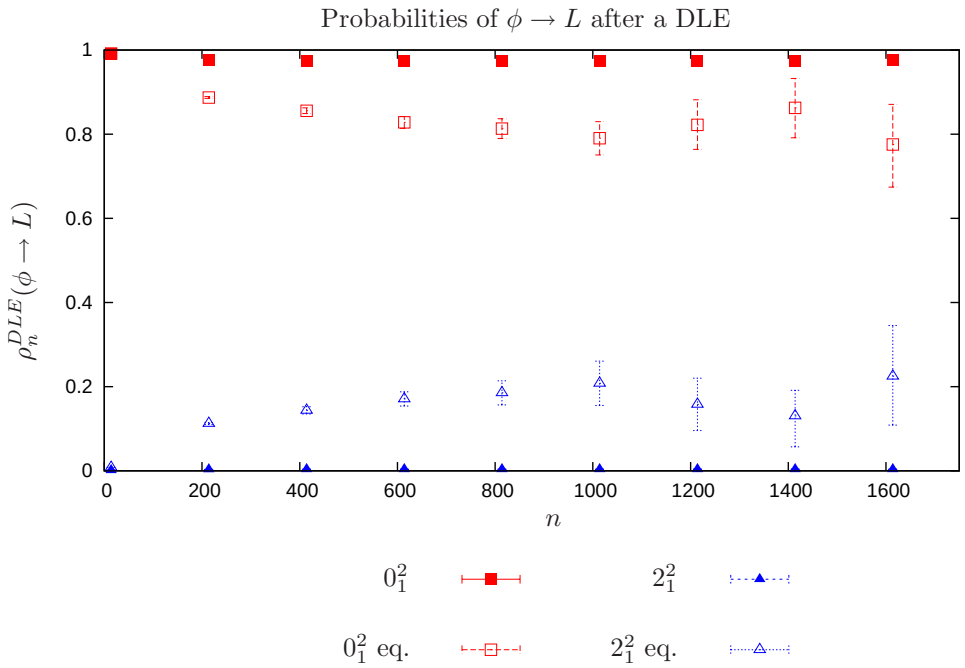


Figure 5.2: A graph of the estimated probability of $\phi \to L$ after a DLE occurs, $\rho_n^{DLE}(\phi \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

result of a recombinase move that forms a link (*i.e.* DLE). Figure 5.3 shows the DLE probability estimates when starting with the positive trefoil $K = 3_1^+$. This graph shows the probabilities of a Hopf link (2_1^2), 4_1^2 and $u(3_1^+ \cup \phi)$ for variable-length polygons (solid symbols) and 'equal-length' polygons (open symbols). For variable-length polygons $u(3_1^+ \cup \phi)$ dominates the probability estimates, however, when looking at 'equal-length' polygons $u(3_1^+ \cup \phi)$ significantly decreases and 2_1^2 and 4_1^2 each increase. Figure 5.4 then shows the corresponding graph for $K = 3_1^-$. Again, in this graph $u(3_1^+ \cup \phi)$ dominates for variable-length polygons, but decreases significantly for 'equal-length' polygons. When starting with $K = 3_1^-$, a 4_1^2 result is very rare so is not shown. In order to compare trends from the LSE model to the tangle analysis in Figure 5.1, both chiralities of trefoils need to be combined in order to see all results of a recombinase move on a chiral knot K. Figure 5.5 shows the probability estimates when both trefoils are combined. From this graph it can be seen that for 'equal-length' polygons it is possible to move 'up' the unlinking pathway to 4_1^2, however the most likely link after a DLE is the Hopf link, 2_1^2 which corresponds to moving 'down' the pathway.

Although the figure eight, $K = 4_1$, does not appear on the unlinking pathway in Figure 5.1, the most likely link after a DLE on $K = 4_1$ should still be a step down in complexity. $K = 4_1$ is an achiral knot, therefore Figure 5.6 shows the graph of all the link transition probability estimates for variable-length polygons and 'equal-length' polygons. The link, 5_1^2, corresponding to 'up' the unlinking pathway isn't shown on Figure 5.6 since the probability of such a link is very low for both variable-length and 'equal-length' polygons. For 'equal-length' polygons it can be seen that the link is most likely simplified to 2_1^2 when a single DLE move is

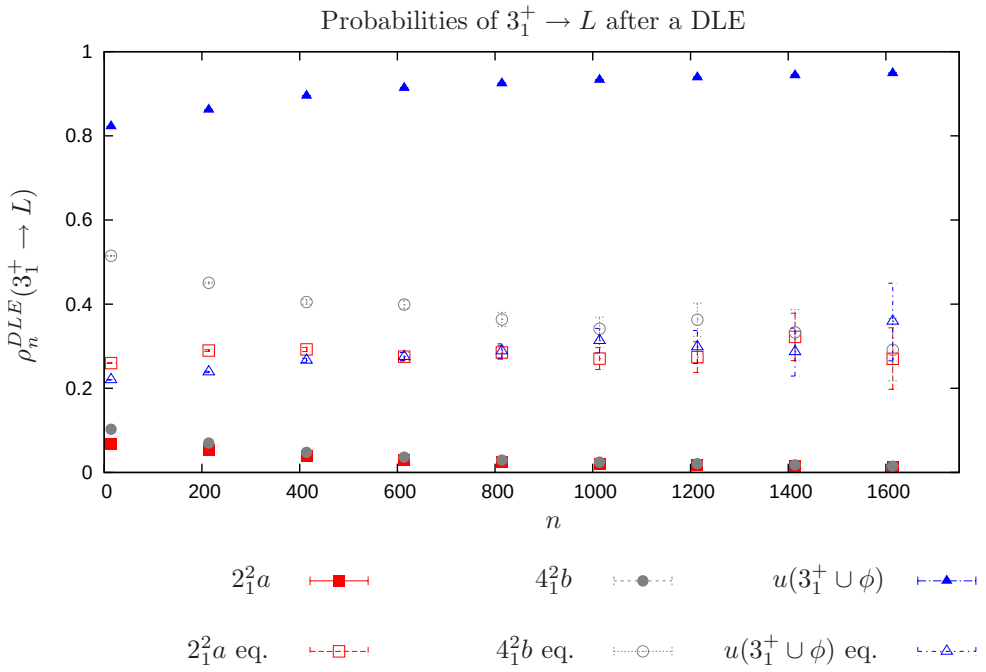


Figure 5.3: A graph of the estimated probability of $3_1^+ \to L$ after a DLE occurs, $\rho_n^{DLE}(3_1^+ \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

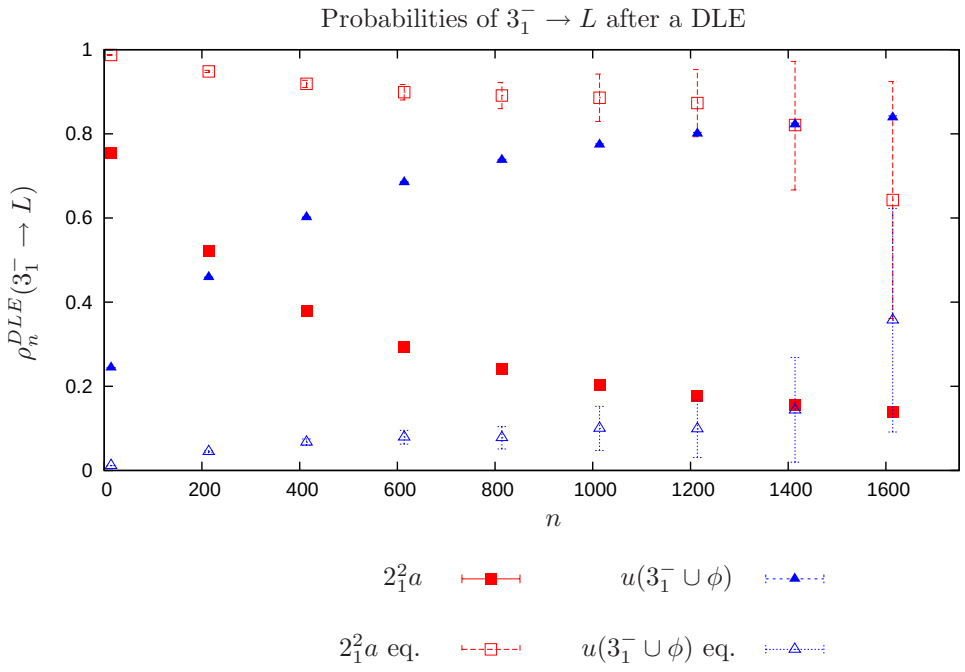


Figure 5.4: A graph of the estimated probability of $3_1^- \to L$ after a DLE occurs, $\rho_n^{DLE}(3_1^- \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

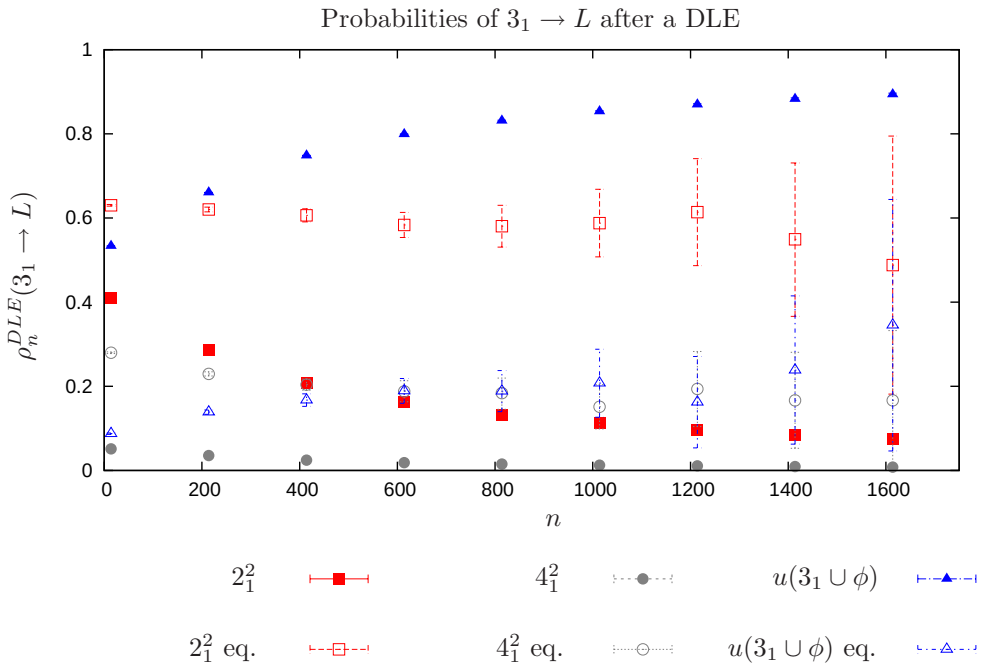


Figure 5.5: A graph of the estimated probability of $3_1 \to L$ after a DLE occurs, $\rho_n^{DLE}(3_1 \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

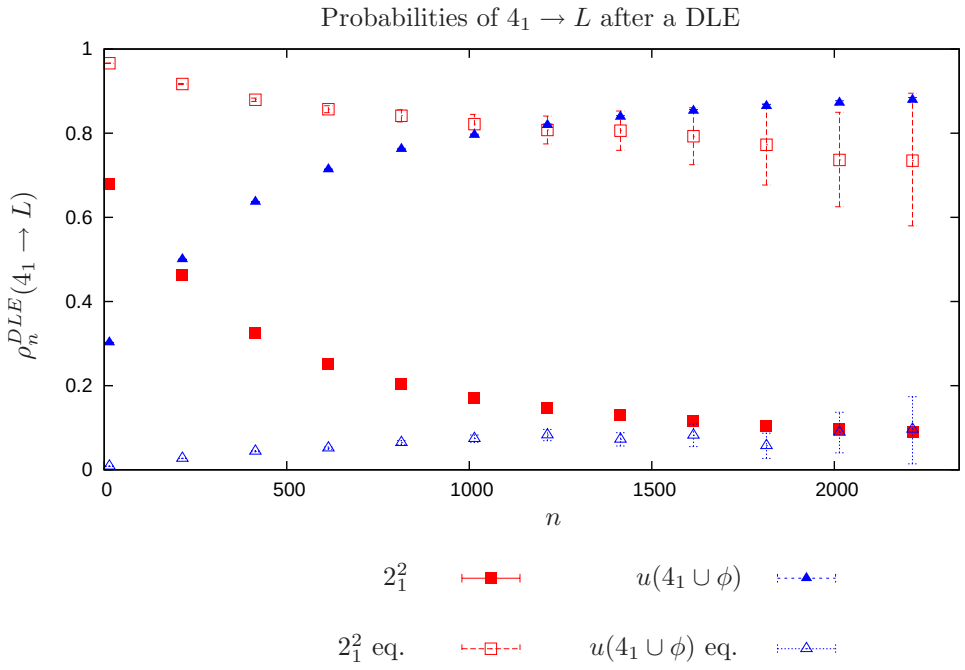


Figure 5.6: A graph of the estimated probability of $4_1 \to L$ after a DLE occurs, $\rho_n^{DLE}(4_1 \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

performed.

Figure 5.1 (b) doesn't specifically show $K = 5_1$, but when starting with this positive five-crossing knot, the unlinking path in Figure 5.1 (a) says that 4_1^2 should be the result after a recombinase move that forms a link (*i.e.* DLE). Like with the trefoil, Figure 5.7 shows the probability estimates associated with $K = 5_1^+$, Figure 5.8 shows the probability estimates associated with $K = 5_1^-$, and Figure 5.9 shows the probability estimates for the combination of both 5_1's. Like with the trefoil, when both chiralities are combined together the full recombinase picture can be seen. Again, when a DLE is performed the most probable link, 4_1^2, corresponds to moving 'down' the unlinking pathway.

And finally, Figure 5.10 shows the probability estimates for $K = 5_2^+$, Figure 5.11 shows the probability estimates for $K = 5_2^-$ and Figure 5.12 shows the combined probability estimates. One last time, it can be seen that $K = 5_2$ obeys the following trend: when studying 'equal-length' polygons the dominant link type after a DLE is a step 'down' the unlinking pathway, here 4_1^2.

As stated earlier, there is limited data from the LSE model for the subset of polygons where the SAWs are exactly the same length, or even within 24-edges of each other, so no formal analysis was done; however, the preliminary results presented here are very promising. Starting with all knot types, when looking at 'equal-length' polygons the link transition probability estimates after a DLE are dominated by the link which is considered simpler, or "down" the unlinking path. This result agrees with the results from tangle analysis

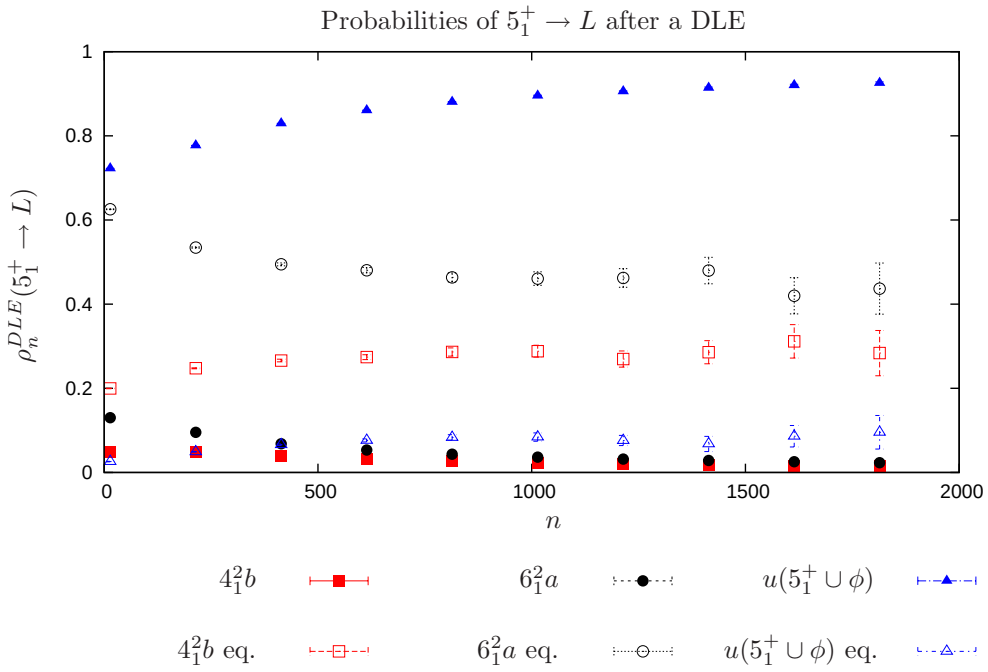


Figure 5.7: A graph of the estimated probability of $5_1^+ \to L$ after a DLE occurs, $\rho_n^{DLE}(5_1^+ \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

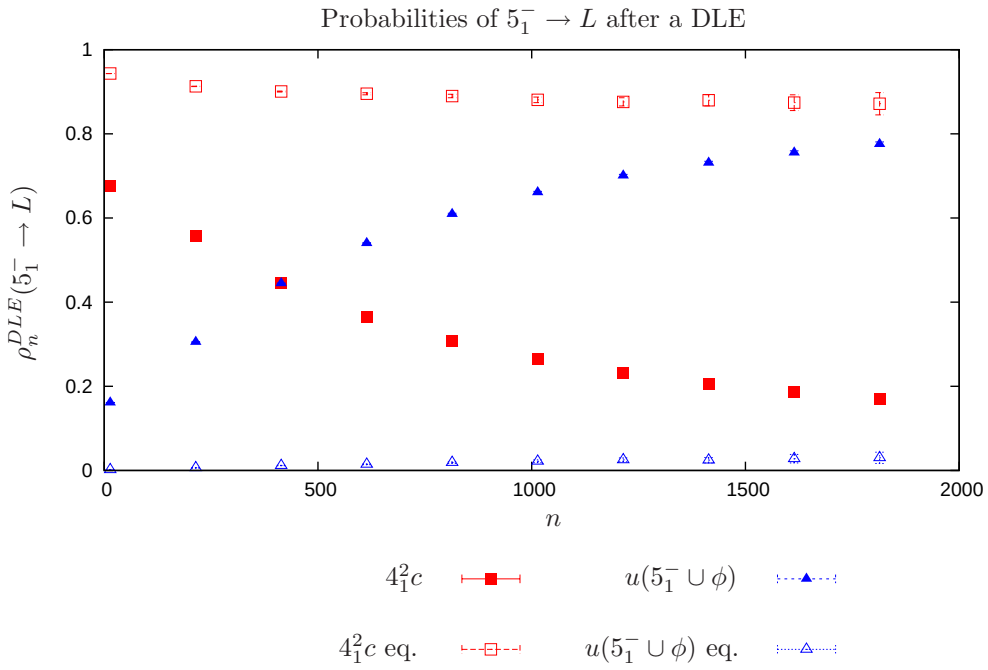


Figure 5.8: A graph of the estimated probability of $5_1^- \to L$ after a DLE occurs, $\rho_n^{DLE}(5_1^- \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

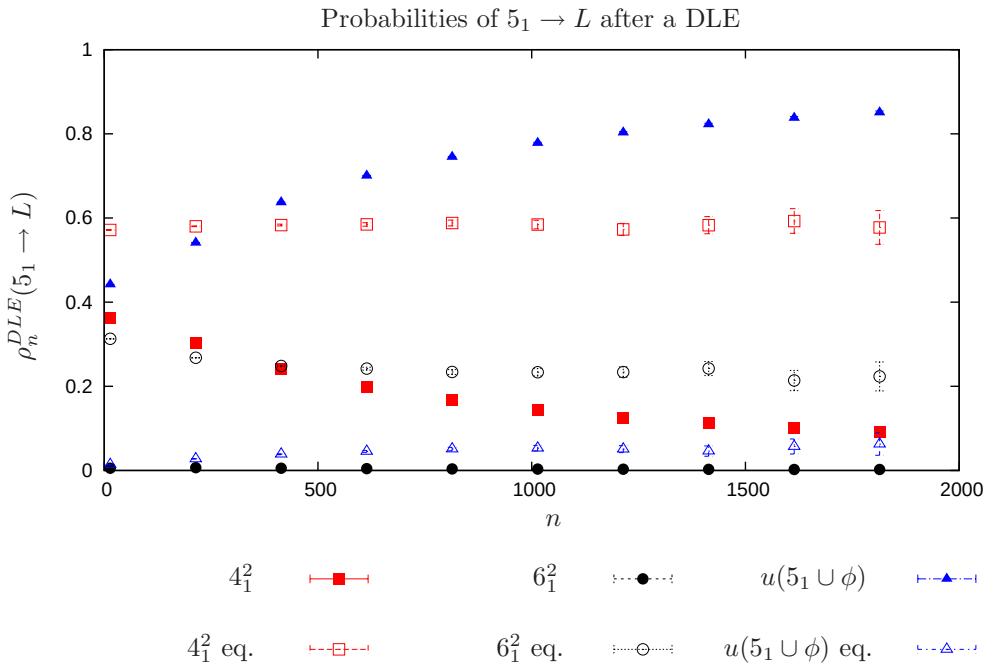


Figure 5.9: A graph of the estimated probability of $5_1 \to L$ after a DLE occurs, $\rho_n^{DLE}(5_1 \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

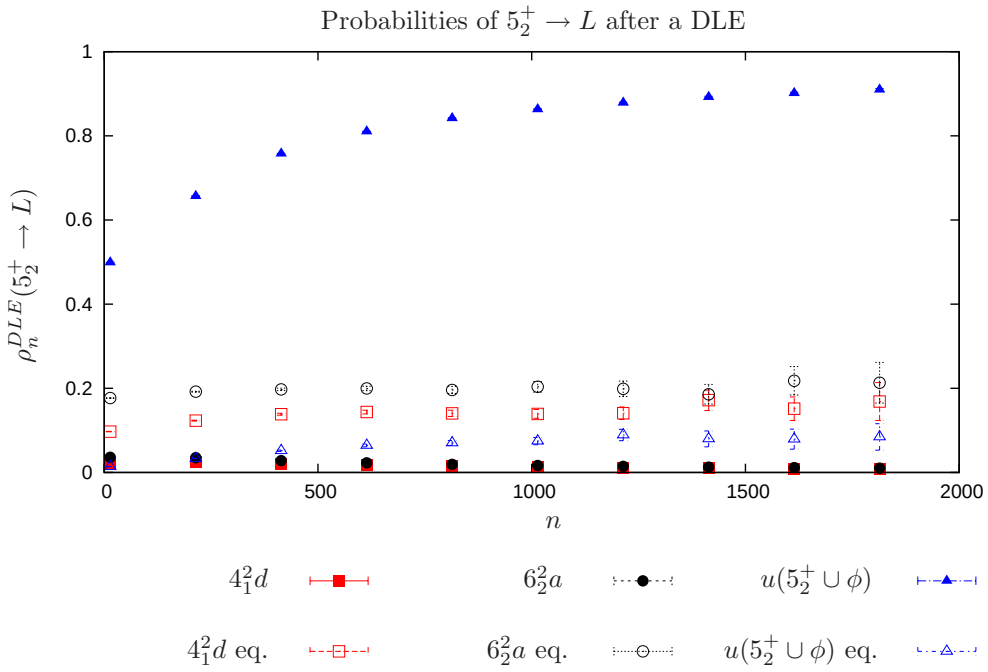


Figure 5.10: A graph of the estimated probability of $5_2^+ \to L$ after a DLE occurs, $\rho_n^{DLE}(5_2^+ \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

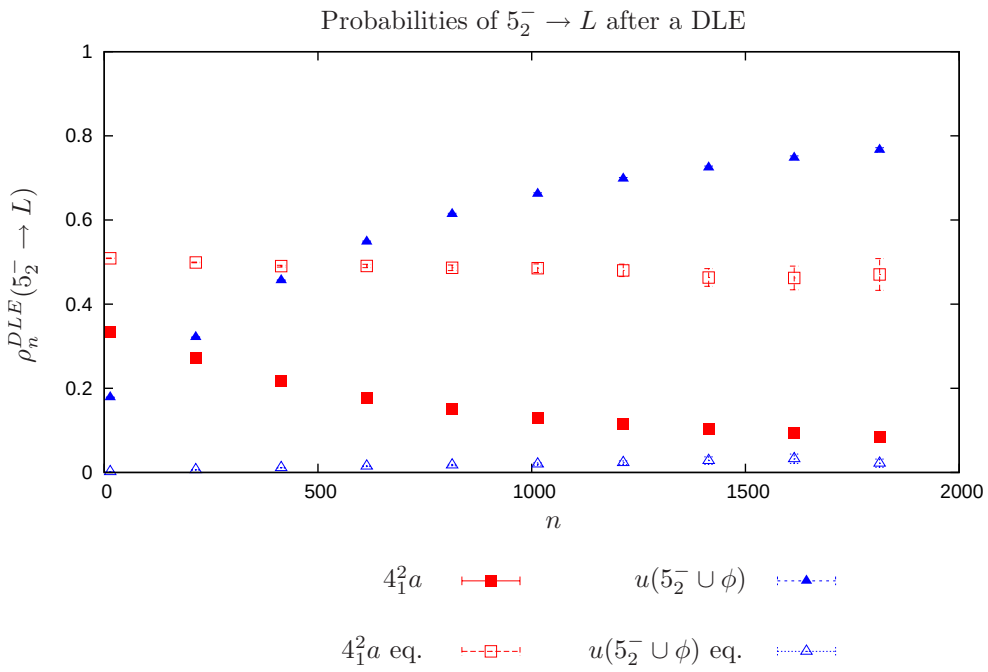


Figure 5.11: A graph of the estimated probability of $5_2^- \to L$ after a DLE occurs, $\rho_n^{DLE}(5_2^- \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

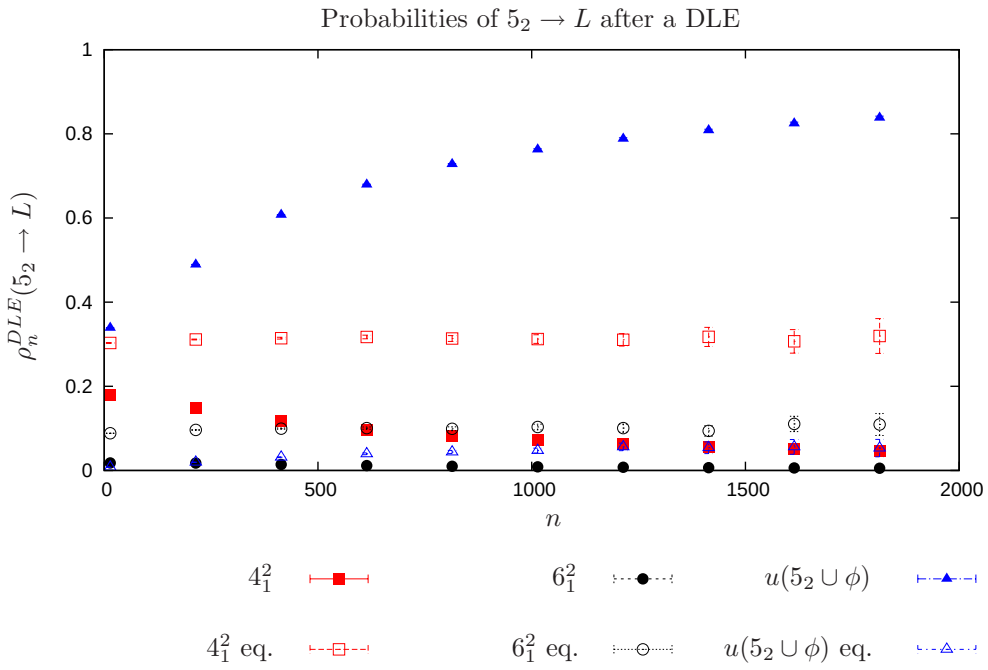


Figure 5.12: A graph of the estimated probability of $5_2 \to L$ after a DLE occurs, $\rho_n^{DLE}(5_2 \to L)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

and also has some biological significance. These results show that when the polygons relate closest to what DNA would look like, namely when the *dif* sites are mid-cell and split the dimer in half, the result of a modelled recombinase move simplifies the knot. This agrees with what both Ip *et al.* and Grainge *et al.* saw in their experiments; the unlinking pathway is step-wise and simplifies the knot gradually [47, 53].

5.3 IKE 'Equal-Length' Polygon Results

Although the unlinking pathway (Figure 5.1 (a)) is only looking at the path from link to knot to link, the other possibility is that an IKE occurs, going from knot to knot. In this case it is of interest to see if the starting knot type (K) is simplified after an IKE or made more complex. Therefore, for each starting knot type K, the following estimates are obtained: the link transition probability estimates of going to a less complex knot (a lower minimal crossing number), the link transition probability estimate of going to a more complex knot (a higher minimal crossing number), and the link transition probability estimate of staying the same knot (the same minimal crossing number). Again these probabilities are investigated for variable-length polygons and 'equal-length' polygons. For these results, the term less complex, or 'simple', refers to all knots with a lower crossing number and the term more complex, or 'complex' refers to knot types that have a higher crossing number. For starting knot type ϕ, there is no simpler knot. For some knot types the only simpler knot is ϕ, so it will be labelled as such. There is often more than three resulting knots which are more complex, however only at most three probabilities will be included for convenience (less than three is used when one or two knot types dominate). The results for these estimates are shown in Figures 5.13, 5.14, 5.15, 5.16, and 5.17 for each starting knot type $K = \phi, 3_1, 4_1, 5_1, 5_2$, respectively.

Figures 5.13 shows that staying the unknot has the greatest probability as there is no less complex knot. For all the other graphs it can be seen that for variable-length polygons the most probable knot is the starting knot, however, when looking at 'equal-length' polygons, the most probable knot is the less complex, or simpler knot in all cases. These results are encouraging when looking from a biological perspective, as it's been shown that recombinase is capable of acting like topoisomerase, meaning it can be efficient at unknotting DNA. These modelled results support that recombinase is capable of unknotting polygons efficiently when 'equal-length' polygons are considered and when the overall length of the polygon is not too large.

5.4 Chapter 5 Summary

The results shown above for the subset of Θ-SAPs which are 'equal-length' are very promising. Although lack of data has prevented full analysis (see Section 6.2 for more on this), these preliminary results show that the LSE model simplifies the starting knot type K in both the DLE and the IKE cases. For the DLE case, the unlinking pathway shown in Figure 5.1 (a) was shown to be the most probable outcome of a single recombinase move, regardless of the initial knot type K. The entire unlinking pathway cannot be investigated as only one recombinase move is performed using the LSE, and the starting polygons for the LSE model are

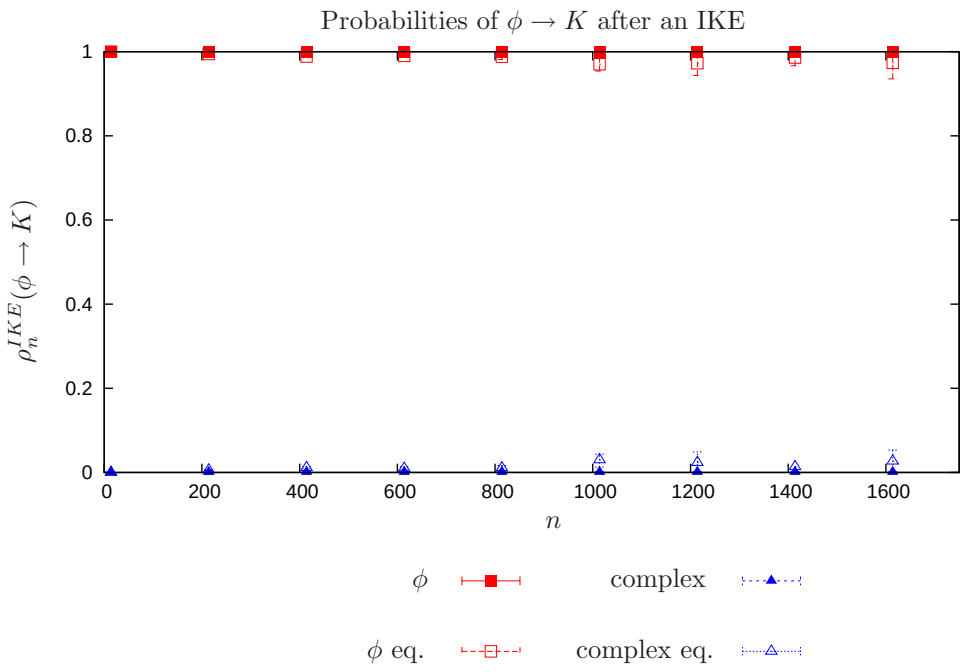


Figure 5.13: A graph of the estimated probability of $\phi \to K$ after an IKE occurs, $\rho_n^{IKE}(\phi \to K)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

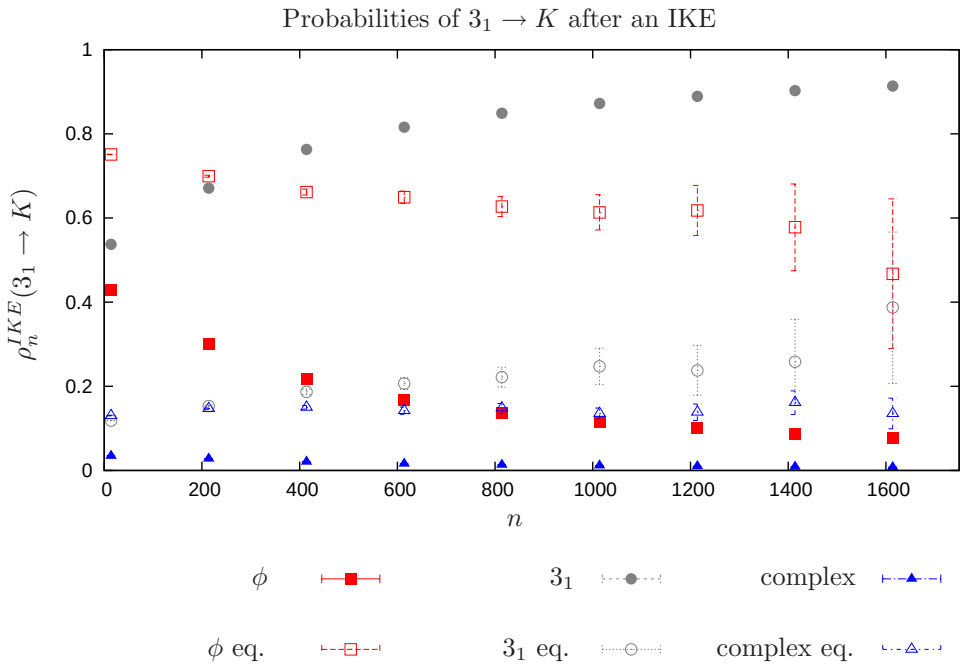


Figure 5.14: A graph of the estimated probability of $3_1 \to K$ after an IKE occurs, $\rho_n^{IKE}(3_1 \to K)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

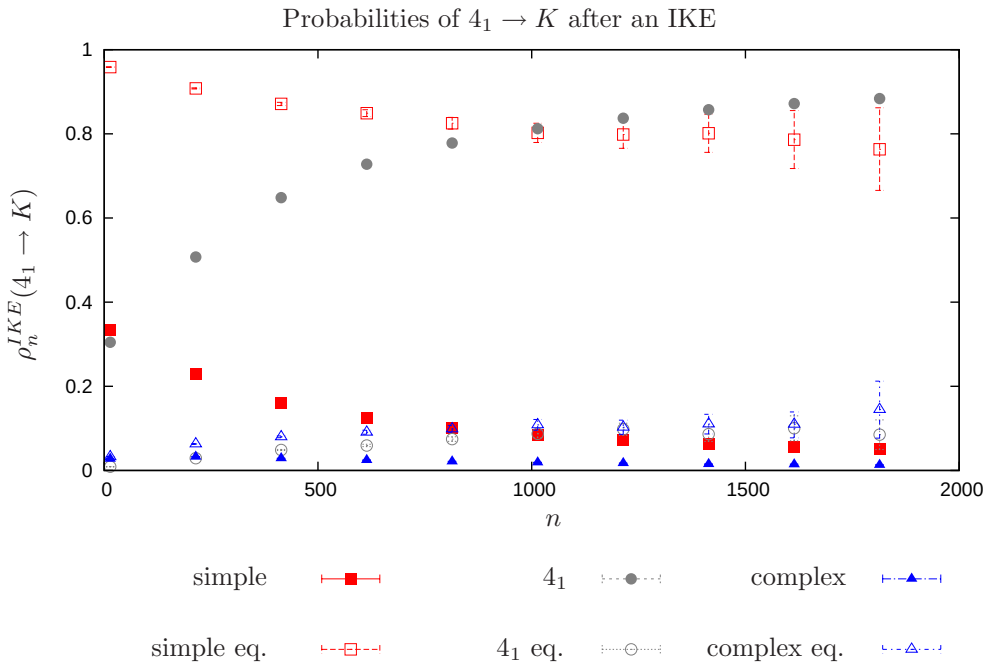


Figure 5.15: A graph of the estimated probability of $4_1 \to K$ after an IKE occurs, $\rho_n^{IKE}(4_1 \to K)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

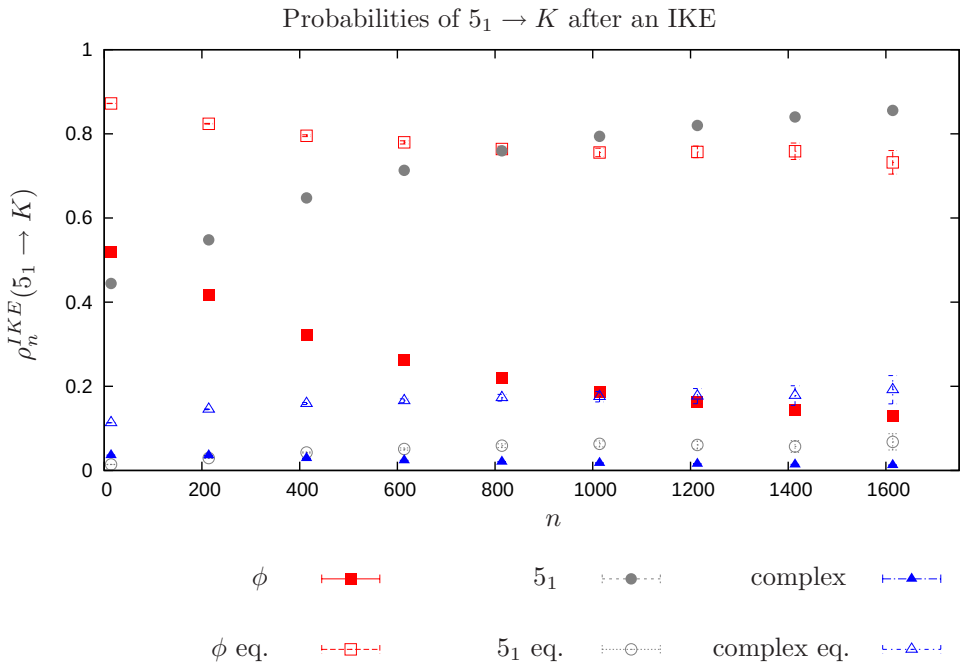


Figure 5.16: A graph of the estimated probability of $5_1 \to K$ after an IKE occurs, $\rho_n^{IKE}(5_1 \to K)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

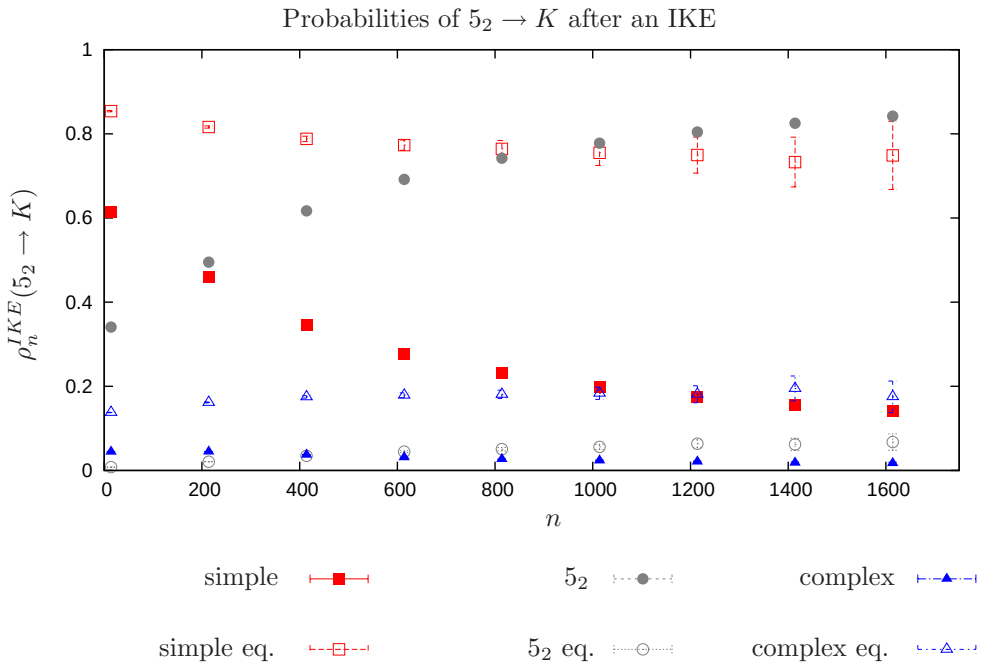


Figure 5.17: A graph of the estimated probability of $5_2 \to K$ after an IKE occurs, $\rho_n^{IKE}(5_2 \to K)$. Probabilities are shown for variable-length polygons (solid symbols) and for 'equal-length' polygons (open symbols). Each data point is plotted with error bars representing 95% confidence intervals.

only knots (see Section 6.2 for more on this). For the IKE case, it was shown that when looking at Θ-SAPs which are 'equal-length', the most probable outcome is for the starting knot to be simplified after an IKE (aside from ϕ which cannot be simplified). Both of these results are very encouraging and lend support to more research being performed with the LSE model (see Section 6.2 for more on this). Note that these results are limited to polygons of length less than $\approx$1600, *i.e.* each side is $\approx$800 edges. Polymer scaling theory suggests that as lengths get larger these trends will change since the *Case 2* scenario will start to dominate. However, these smaller length polygons are expected to be relevant to sizes in DNA experiments.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis began with a review of the enzymes topoisomerase (*cf.* Section 1.1.2) and recombinase (*cf.* Section 1.1.3). It was shown that both these enzymes are capable of unknotting and unlinking DNA, via a strand passage and a strand exchange, respectively. These unknotting and unlinking abilities allow all cells to replicate and survive. An overview of previous enzyme-DNA models (*cf.* Section 1.3) was represented, and it was established that link transition probabilities, the probabilities of going from knot type K to link type L after a strand exchange, have yet to be investigated for recombinase models.

It was introduced that self-avoiding polygons (SAPs) on the simple cubic lattice can represent DNA configurations (*cf.* Section 2.1). A review was then presented of the Local Strand Passage (LSP) model developed by Szafron and Soteros [106, 107] to model topoisomerase-DNA interactions with a fixed structure, called Θ, at the interaction location (*cf.* Section 2.2). Following this, the new Local Strand Exchange (LSE) model was introduced for modelling the recombinase-DNA interaction, again with the Θ structure at the interaction location (*cf.* Section 2.3). Recombinases's action was then modelled with two different structures to take into account the two biological types of recombination: 1) direct repeat recombination sites were modelled with the direct-repeat-to-link strand exchange (DLE), and 2) inverted repeat recombination sites were modelled with the inverted-repeat-to-knot strand exchange (IKE).

The new LSE model was applied to previously generated CMC Θ-BFACF simulation data (*cf.* Section 3.4) in order to investigate link transition probabilities. A list of all observed knot and link types after a strand exchange was compiled (*cf.* Appendix B). Strong numerical evidence was provided showing that the link transition probabilities from the newly created LSE model have asymptotic behaviours which satisfy conjectures from polymer scaling theory (*cf.* Section 4.2). In particular, it was shown that as the lengths of Θ-SAPs increase towards infinity, it becomes less likely that the fixed structure Θ will interact with the pre-existing knot. This means that the original knot remains intact and a secondary knot or link can be formed at Θ. In the DLE case, a strand exchange results in a link with one component of the link having the original knot in its knot decomposition, and, in the IKE case, a composite knot can be created, with one of the knot's components containing the same knot type as the original knot. The LSE model was then compared to the LSP model, highlighting trends across the two different models.

Numerical evidence was also provided showing for a subset of Θ-SAPs with Θ in the middle of the polygons, that it is most probable that the complexity of the knot is reduced after a strand exchange (*cf.* Section 5). In the DLE case, it is most probable that the resulting link will follow the mathematical and biological unlinking pathway. In the IKE case, it is most probable that the knot will be simplified, signifying recombinases's unknotting abilities.

The LSE model has been shown to support previous research, while also providing a good first-step towards modelling enzyme-DNA interactions, specifically recombinase-DNA interactions.

6.2 Potential Future Work

Since the Local Strand Exchange (LSE) model is a first-step model, there is great potential for future work.

An important aspect of recombination that was not accounted for in this model is recombinases's iterative action. Biologically, recombination occurs in a step-wise process, taking a link to a knot to a link in order to finally result in two separate, unlinked monomers. The LSE model currently performs a single recombination move on a knotted Θ-SAP. The next step would be to allow links as the starting state in the LSE model, in order to investigate the link-to-knot transition probabilities, and also model the case when the recombination sites are in separate molecules. If multiple recombinase steps are to be modelled at one time, it is likely that a 'reset' stage will be needed (such as in the tangle analysis shown in Figure 2.10). To incorporate this, secondary DLE and IKE structures may be needed.

An additional next step, with the potential to make the LSE model more biologically realistic, would be to add quantities such as the Yukawa potential (similar to Schmirler's work [89]) to take salt concentrations into account, or to use more complicated lattices, such as the face-centered cubic lattice or the body-centered cubic lattice, to explore which features are lattice-independent. Additionally, due to the geometrical layout of Θ, the recombination sites in the LSE model are not geometrically parallel as in other recombinase models. Therefore, a new fixed structure could be determined which would allow the recombination sites to be geometrically parallel. This would also allow for the study of new structures, since it has been shown that the specific choice of structure does affect the link transition probabilities.

In Chapter 4 it was stated that some fits for IKE *Case 2* were quite poor (as seen in Table 4.8). The main statistical package used in this work was R, however fits were also attempted using the computing environment MATLAB. Neither of these packages were able to achieve acceptable fits, however it is possible that by using other statistical packages, fits could be achieved. In Chapter 5 it was stated that not enough data was generated in order to do a full analysis of polygons with two equal length SAWs. With more generated data this would be possible, allowing for a better comparison to the results from tangle analysis, as well as biological experiments. A new MCMC simulation is currently being run in order to generate four times as many polygons. It is expected that with this new data, a full analysis of polygons with two equal SAWs will be able to be performed.

Another area of interest is to make a more extensive comparison of the LSP and LSE models. In this thesis only a brief comparison was performed, however more work is expected to lead to interesting results. For example, by comparing the knot reduction factors, the unknot/knot ratio could be investigated for the two models to see which model is more efficient at knot reduction. Also, in Chapter 4 it was mentioned that the LSP model has been applied to a second fixed structure called the Symmetric Structure. An equivalent strand exchange could be constructed for this structure and the LSE model could be applied in order to see the changes in the link transition probabilities.

The LSE model was created to model recombinase-DNA interactions, however, it provides a general model for any local enzyme-DNA interaction. For this reason it would be of interest to have a user-friendly tool in which these enzyme-DNA interactions could be viewed and manipulated. This self-contained tool would allow the user to choose the specific enzyme of interest, and then visualize a model of the selected enzyme acting on the DNA. The user would be able to view statistical values, such as link transition probabilities, or view the action in mathematical terms such as via tangle analysis tools. This type of tool would allow students and researchers to be introduced to this area of research much more easily, as well as provide applications for biologists to explore specific analysis or visual questions.

Finally, the tables found in Appendix B list all the observed link transitions for the LSE model. It is expected that other research groups will be interested in these observations, for example a tangle analysis expert could fully characterize the DLE and IKE moves from the observed link transitions. Therefore, a public website with this information and larger versions of the figures shown in Appendix B is expected to be useful to the research community in this area.

References

[1] Colin C. Adams. *The Knot Book*. American Mathematical Society, New York, 1994.

[2] Toshiwo Andoh and Ryoji Ishida. Catalytic inhibitors of DNA topoisomerase II. *Biochimica et Biophysica - Gene Structure and Expression*, 1400(1-3):155–171, 10 1998.

[3] Javier Arsuaga, Mariel Vazquez, Sonia Trigueros, and Joaquim Roca. Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proceedings of the National Academy of Sciences*, 99(8):5373–5377, 4 2002.

[4] M. Atapour, C.E. Soteros, and S.G. Whittington. Stretched polygons in a lattice tube. *Journal of Physics A: Mathematical and Theoretical*, 42(32):322002–322011, 7 2009.

[5] Laurent Aussel, François-Xavier Barre, Mira Aroyo, Andrzej Stasiak, Alicja Z. Stasiak, and David Sherratt. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell*, 108(2):195–205, 2 2002.

[6] M. Baiesi and E. Orlandini. Universal properties of knotted polymer rings. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 86(3):031805, 9 2012.

[7] M. Baiesi, E. Orlandini, and A.L. Stella. The entropic cost to tie a knot. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(6):P06012, 6 2010.

[8] François-Xavier Barre, Mira Aroyo, Sean D. Colloms, Annett Helfrich, François Cornet, and David J. Sherratt. FtsK functions in the processing of a Holliday junction intermediate during bacterial chromosome segregation. *Genes and Development*, 15(5):1172–1181, 4 2000.

[9] François-Xavier Barre and David J. Sherratt. Chromosome dimer resolution. In N. Patrick Higgins, editor, *The Bacterial Chromosome*, chapter 28, pages 513–524. ASM Press, Washington, DC, 2005.

[10] François-Xavier Barre, Britta Søballe, Bénédicte Michel, Mira Aroyo, Malcolm Robertson, and David Sherratt. Circles: The replication-recombination-chromosome segregation connection. *Proceedings of the National Academy of Sciences*, 98(15):8189–8195, 7 2001.

[11] Andrew D. Bates, James M. Berger, and Anthony Maxwell. The ancestral role of ATP hydrolysis in type II topoisomerases: Prevention of DNA double-strand breaks. *Nucleic Acids Research*, 39(15):6327–6339, 4 2011.

[12] Andrew D. Bates and Anthony Maxwell. *DNA Topology*. Oxford University Press, New York, 2005.

[13] B. Berg and D. Foester. Random paths and random surfaces on a digital computer. *Physics Letters B.*, 106(41):323–326, 11 1981.

[14] James M. Berger, Steven J. Gamblin, Stephen C. Harrison, and James C. Wang. Structure and mechanism of DNA topoisomerase II. *Nature*, 379(6562):225–232, 1 1996.

[15] Garry Blakely, Gergard May, Richard McCulloch, Lidia K. Arciszewska, Mary Burke, Susan T. Lovett, and David J. Sherratt. Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell*, 75(2):351–361, 10 1993.

[16] Dorothy Buck. DNA topology. In Dorothy Buck and Erica Flapan, editors, *Proceedings of Symposia in Applied Mathematics: Applications of Knot Theory*, volume 66, chapter 2, pages 47–79. American Mathematical Society, Providence, Rhode Island, 2009.

[17] Y. Burnier, C. Weber, A. Flammini, and A. Stasiak. Local selection rules that can determine specific pathways of DNA unknotting by type II DNA topoisomerases. *Nucleic Acids Research*, 35(15):5223–5231, 8 2007.

[18] Aragão De Carvalho and S. Caracciolo. A new Monte Carlo approach to the critical properties of self-avoiding random walks. *Journal de Physique*, 44(3):323–331, 1983.

[19] Aragão De Carvalho, S. Caracciolo, and J. Fröhlich. Polymers and $g|\phi|^4$ theory in four dimensions. *Nuclear Physics Section B*, 215(2):209–248, 1983.

[20] Marla A. Cheston, Kevin McGregor, Christine E. Soteros, and Michael L. Szafron. New evidence on the asymptotics of knotted lattice polygons via local strand-passage models. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(2):P02014, 2 2014.

[21] Nathan Clisby, Richard Liang, and Gordon Slade. Self-avoiding walk enumeration via the lace expansion. *The Journal of Physics A: Mathematical and Theoretical*, 40(36):10973–11017, 8 2007.

[22] Sean D. Colloms. The topology of plasmid-monomerizing Xer site-specific recombination. *Biochemical Society Transactions*, 41(2):589–594, 4 2013.

[23] Sean D. Colloms, Richard McCulloch, Karen Grant, Liam Neilson, and David J. Sherratt. Xer-medicated site-specific recombination *in vitro*. *The EMBO Journal*, 15(6):1172–1181, 4 1996.

[24] Peter Cromwell. *Knots and Links*. Cambridge University Press, United Kingdom, 2004.

[25] Isabel K. Darcy. Biological distances on DNA knots and links: Applications to Xer recombination. *Journal of Knot Theory and Its Ramifications*, 10(2):269–294, 4 2001.

[26] Isabel K. Darcy and Robert G. Scharein. TopoICE-R: 3D visualization modeling the topology of DNA recombination. *Bioinformatics Applications Note*, 22(14):1790–1791, 5 2006.

[27] Frank B. Dean, Andrezej Stasiak, Theo Koller, and Nicholas R. Cozzarelli. Duplex DNA knots produced by *Escherichia coli* topoisomerase I. *The Journal of Biological Chemistry*, 260(8):4975–4983, 4 1985.

[28] Richard W. Deibler, Jennifer K. Mann, De Witt L. Sumners, and Lynn Zechiedrich. Hin-mediated DNA knotting and recombining promote replicon dysfunction and mutation. *BMC Molecular Biology*, 8(44), 5 2007.

[29] Cheikh Tidiane Diagne, Maya Salhi, Estell Crozat, Lauren Salome, Francois Cronet, Philippe Rousseau, and Catherine Tardin. TPM analyses reveal that FtsK contributes both to the assembly and the activation of the XerCD-*dif* recombination synapse. *Nucleic Acids Research*, 42(3):1721–1732, 11 2014.

[30] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Society, Rhode Island, 2009.

[31] Genetics Education. DNA double helix.
`https://www.flickr.com/photos/119980645@N06/13081113544//`, 4 2014. CC BY 2.0.

[32] Genetics Education. Separation of DNA base pairs.
`https://www.flickr.com/photos/119980645@N06/13080775743/`, 4 2014. CC BY 2.0.

[33] Jeremy Eng. Self-avoiding polygons in (l,m)-tubes. Master's thesis, University of Saskatchewan, 9 2014.

[34] C. Ernst and D.W. Sumners. A calculus for rational tangles: Applications to DNA recombination. *Mathematical Proceedings of the Cambridge Philosophical Society*, 108(3):489–515, 11 1990.

[35] Olivier Espeli, Chong Lee, and Kenneth Marians. A physical and functional interaction between *Escherichia coli* and FtsK and topoisomerase iv. *The Journal of Biological Chemistry*, 278(45):44639–44644, 11 2003.

[36] C.M. Field and D.K. Summers. Multicopy plasmid stability: Revisiting the dimer catastrophe. *Journal of Theoretical Biology*, 291:119–127, 12 2011.

[37] George S. Fishman. *Monte Carlo: Concepts, algorithms, and applications*. Springer, New York, 1997.

[38] Alessandro Flammini, Amos Maritan, and Andrzej Stasiak. Simulations of action of DNA topoisomerases to investigate boundaries and shapes of spaces of knots. *Biophysical Journal*, 87(5):2968–2975, 11 2004.

[39] Jonathan M. Fogg, Daniel J. Catanese Jr., Graham L. Randall, Michell C. Seick, and Lynn Zechiedrich. Differences between positively and negatively supercoiled DNA that topoisomerases may distinguish. In Craig John Benham, Stephen Harvey, Wilma K. Olson, De Witt L. Sumners, and David Swigon, editors, *Mathematics of DNA Structure, Function and Interactions*, volume 150, chapter 5, pages 73–122. Springer, New York, 2009.

[40] Rosalind E. Franklin and R.G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740–741, 4 1953.

[41] P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millett, and A. Ocneanu. New polynomial invariant of knots and links. *Bulletin of the American Mathematical Society*, 12(2):239–247, 4 1985.

[42] Stacie J. Froelich-Ammon and Neil Osheroff. Topoisomerase poisons: Harnessing the dark side of enzyme mechanism. *The Journal of Biological Chemistry*, 270(37):21429–21432, 9 1995.

[43] Andre Gelman. Inference and monitoring convergence. In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, chapter 8, pages 131–143. Champman and Hall, New York, 1996.

[44] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 4 1992.

[45] Ian Grainge. FtsK - A bacterial cell division checkpoint? *Molecular Microbiology*, 78(5):1055–1057, 10 2010.

[46] Ian Grainge. Simple topology: FtsK-directed recombination at the *dif* site. *Biochemical Society Transactions*, 41(2):595–600, 2013.

[47] Ian Grainge, Migena Bregu, Mariel Vazquez, Viknesh Sivanathan, Stephen C.Y. Ip, and David J. Sherratt. Unlinking chromosome catenanes *in vivo* by site-specific recombination. *The EMBO Journal*, 26(19):4228–4238, 9 2007.

[48] Ian Grainge, Christian Lesterlin, and David J. Sherratt. Activation of XerCD-*dif* recombination by the FtsK DNA translocase. *Nucleic Acids Research*, 39(12):5140–5148, 3 2011.

[49] Nigel D.F. Grindley, Katrine L. Whiteson, and Pheobe A. Rice. Mechanisms of site-specific recombination. *Annual Review of Biochemistry*, 75:567–605, 7 2006.

[50] J.M. Hammersley. On the rate of convergence to the connective constant of the hypercubical lattice. *The Quarterly Journal of Mathematics*, 12(1):250–256, 1961.

[51] J.M Hammersley and K.W. Morton. Poor man's Monte Carlo. *Journal of the Royal Statistical Society B: Methodological*, 16(1):23–28, 1954.

[52] Andrias Hojgaard, Heather Szerlong, Camille Tabor, and Peter Kuempel. Norfloxacin-induced DNA cleavage occurs at the *dif* resolvase locus in *Escherichia coli* and is the result of interaction with topoisomerase IV. *Molecular Microbiology*, 33(5):1027–1036, 9 1999.

[53] Stephen C.Y. Ip, Migena Bregu, François-Xavier Barre, and David J. Sherratt. Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *The EMBO Journal*, 22(23):6399–6407, 12 2003.

[54] V.F.R. Jones. Hecke algebra representations of braid groups and link polynomials. *Annals of Mathematics*, 126(2):335–388, 9 1987.

[55] Yacov Kantor. Knots in polymers. *Pramana Journal of Physics*, 64(6):1011–1017, 6 2005.

[56] Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, Inc., New York, 2 edition, 1975.

[57] Junichi Kato, Yukinobu Nishimura, Ryu Inamura, Hironori Niki, Sota Hiraga, and Hideho Suzuki. New topoisomerase essential for chromosome segregation in e. coli. *Cell*, 63(2):393–404, 10 1990.

[58] Mark A. Krasnow, Andrzej Stasiak, Sylvia J. Spengler, Frank Dean, Theo Koller, and Nicholas R. Cozzarelli. Determination of the absolute handedness of knots and catenanes of DNA. *Nature*, 304(5926):559–560, 8 1983.

[59] Nicholas R. Leslie and David J. Sherratt. Site-specific recombination in the replication terminus region of *Escherichia coli*: Functional replacement of *dif*. *The EMBO Journal*, 14(7):1561–1570, 4 1995.

[60] Leroy F. Liu, Richard E. Depew, and James C. Wang. Knotted single-stranded DNA rings: A novel topological isomer of circular single-stranded DNA formed by treatment with *Escherichia coli* ωprotein. *Journal of Molecular Biology*, 106:439–452, 1976.

[61] L.F. Liu, J.L. Davis, and R. Calendar. Novel topologically knotted DNA from bacteriophage p4 capsids: Studies with DNA topoisomerases. *Nucleic Acids Research*, 9(16):3979–3989, 8 1981.

[62] Zhirong Liu and Hue Sun Chan. Efficient chain moves for Monte Carlo simulations of a wormlike DNA model: Excluded volume, supercoils, site juxtapositions, knots and comparisons with random-flight and lattice models. *The Journal of Chemical Physics*, 128(4):145104, 4 2008.

[63] Zhirong Liu, Richard W. Deibler, Hue Sun Chan, and Lynn Zechiedrich. The why and how of DNA unlinking. *Nucleic Acids Research*, 37(3):661–671, 2 2009.

[64] Zhirong Liu, Lynn Zechiedrich, and Hue Sun Chan. Local site preference rationalizes disentangling by DNA topoisomerases. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 81(1):031902, 3 2010.

[65] Neal Madras and Gordon Slade. *The Self-Avoiding Walk*. Springer, New York, 1996.

[66] Maksim. Figure 8 knot 01. `https://commons.wikimedia.org/wiki/File:Figure8knot-01.png`, 4 2006. Public Domain.

[67] Maksim. Trefoil knot 01. `https://commons.wikimedia.org/wiki/File:TrefoilKnot-01.png`, 3 2006. Public Domain.

[68] Magnus Manske. Trefoil knot 02. `https://commons.wikimedia.org/wiki/File:TrefoilKnot-02.png`, 11 2007. Public Domain.

[69] B. Marcone, E. Orlandini, A.L. Stella, and F. Zonta. Size of knots in ring polymers. *Physical Review E: Statistical, Nonlinear and Soft Matter Physics*, 75(4):041105, 4 2007.

[70] Anthony Maxwell and David M. Lawson. The ATP-binding site of type II topoisomerases as a target for antibacterial drugs. *Current Topics in Medicinal Chemistry*, 3(3):283–303, 2003.

[71] Barbara McClintock. A correlation of ring-shaped chromosomes with variegation in *Zea mays*. *Proceedings of the National Academy of Sciences*, 18(12):677–681, 12 1932.

[72] Anthony Montemayor. On nullification of knots and links. Master's thesis, Western Kentucky University, 5 2012.

[73] Hironori Niki, Yoshiharu Yamaichi, and Sota Hiraga. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes and Development*, 14(2):212–223, 1 2000.

[74] E. Orlandini, M.C. Tesi, E.J. Janse van Rensburg, and S.G. Whittington. Entropic exponents of lattice polygons with specified knot type. *Journal of Physics A: Mathematical and General*, 29(12):L299–L303, 6 1996.

[75] E. Orlandini, M.C. Tesi, E.J. Janse van Resburg, and S.G. Whittington. Asymptotics of knotted lattice polygons. *Journal of Physics A: Mathematical and General*, 31(28):5953–5967, 7 1998.

[76] C.W. Patty. *Foundations of Topology*. PWS-Kent Publishing Company, Boston, 1992.

[77] Koryn Pérals, Hervé Capiaux, Jean-Baptiste Vicourt, Jean-Michel Louarn, David J. Sherratt, and François Cornet. Interplay between recombination, cell division and chromosome structure during chromosome dimer resolution in *Escherichia coli*. *Molecular Microbiology*, 39(4):904–913, 2 2001.

[78] Koryn Pérals, François Cornet, Yann Merlet, Isabelle Delon, and Jean-Michel Lourarn. Functional polarization of the *Escherichia coli* chromosome terminus: The *dif* site acts in chromosome dimer resolution only when located between long stretches of opposite polarity. *Molecular Microbiology*, 36(1):33–43, 4 2000.

[79] J. Portillo, Y. Diao, R. Scharein, J. Arsuaga, and M. Vazquez. On the mean and variance of the writhe of random polygons. *Journal of Physics A: Mathematical and Theoretical*, 44(27):275004–275013, 6 2011.

[80] Jerod L. Ptacin, Marcelo Nollmann, Eric C. Becker, Nicholas R. Cozzarelli, Kit Pogliano, and Carlos Bustamante. Sequence-directed DNA export guides chromosome translocation during sporulation in *Bacillu subtilis*. *Nature Structural and Molecular Biology*, 15(5):485–493, 5 2008.

[81] Eric Rawdon, Akos Dobay, John C. Kern, Kenneth C. Millett, Michael Piatek, Patrick Plunkett, and Andrzej Stasiak. Scaling behavior and equilibrium lengths of knotted polymers. *Macromolecules*, 41(12):4444–4451, 5 2008.

[82] Gavin D. Recchia, Mira Aroyo, Daniel Wolf, Garry Blakely, and David J. Sherratt. FtsK-dependent and -independent pathways of Xer site-specific recombination. *The EMBO Journal*, 18(20):5724–5734, 10 1999.

[83] Dale Rolfsen. *Knots and Links*. American Mathematical Society, Providence, Rhode Island, 2003.

[84] Isreal Rosenfield, Edward Ziff, and Borin Van Loon. *DNA for Beginners*. Writers and Readers, London, 1983.

[85] Sheldon M. Ross. *Introduction to Probability Models*. Academic press, USA, 10 edition, 2010.

[86] Valentin V. Rybenkov, Christian Ullsperger, Alexander V. Vologodskii, and Nicholas R. Cozzarelli. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science*, 277(5326):690–693, 8 1997.

[87] Yuki Saka and Mariel Vazquez. TangelSolve: Topological analysis of site-specific recombination. *Bioinformatics Applications Note*, 2002(7):1011–1012, 1 2002.

[88] Robert G. Scharein. KnotPlot. `http://knotplot.com/`, 10 2014.

[89] Matthew Schmirler. Strand passage and knotting probabilities in an interacting self-avoiding polygon model. Master's thesis, University of Saskatchewan, 9 2012.

[90] R.D. Schram, G.T. Barkema, and R.H. Bisseling. Exact enumeration of self-avoiding walks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(06):P06019, 6 2011.

[91] David J. Sherratt. Bacterial chromosome dynamics (review). *Science*, 301(5634):780–785, 8 2003.

[92] David J. Sherratt, Lidia K. Arciszewska, Garry Blakely, Sean Colloms, Karen Grant, Nick Leslie, and Richard McCulloch. Site-specific recombination and circular chromosome segregation. *Philosophical Transactions: Biological Sciences*, 347(1319):37–42, 1 1995.

[93] David J. Sherratt, Ivy F. Lau, and François-Xacier Barre. Chromosome segregation. *Current Opinion in Microbiology*, 4(6):653–659, 12 2001.

[94] David J. Sherratt, Britta Søballe, Fraçois-Xavier Barre, Sergio Filipe, Ivy Lau, Thomas Massey, and James Yates. Recombination and chromosome segregation. *The Royal Society*, 359(1441):61–69, 1 2004.

[95] Koya Shimokawa, Kai Ishihara, Ian Grainge, David J. Sherratt, and Mariel Vazquez. FtsK-dependent XerCD-*dif* recombination unlinks replication catenanes in a stepwise manner. *Proceedings of the National Academy of Sciences*, 110(63):20906–20911, 12 2013.

[96] Alan Sokal and Lawrence Thomas. Exponential convergence to equilibrium for a class of random-walk models. *Journal of Statistical Physics*, 54(3):797–828, 1989.

[97] Alan D. Sokal. Monte Carlo methods in statistical mechanics: Foundations and new algorithms - lectures at cargèse summer school on "functional integration: Basics and applications", 9 1996.

[98] C.E. Soteros. Knots in graphs in subsets of $\mathbb{Z}^3$. *Institute for Mathematics and its Applications*, 103:101–134, 1998.

[99] C.E. Soteros, D.W. Sumners, and S.G. Whittington. Entanglement complexity of graphs in $\mathbb{Z}^3$. *Mathematical Proceedings of the Cambridge Philosophical Society*, 111(1):75–91, 1 1992.

[100] Andrzej Stasiak, Vsevolod Katritch, Jan Bednar, Didier Michoud, and Jacques Dubochet. Eletrophoretic mobility of DNA knots. *Nature*, 384(6605):122, 11 1996.

[101] Walter Steiner, Guowen Liu, William D. Donachie, and Peter Kuempel. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Molecular Microbiology*, 31(2):579–583, 5 1999.

[102] Walter W. Steiner and Peter L. Kuempel. Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the *dif* resolvase site. *Journal of Bacteriology*, 180(23):6269–6275, 12 1998.

[103] D.W. Sumners and S.G. Whittington. Knots in self-avoiding walks. *Journal of Physics A: Mathematical and General*, 21(7):1689–1694, 1988.

[104] Michael Szafron. An IMA mini-course: Statistical methods used in the simulation of knotting and linking of biopolymers, 4 2014.

[105] Michael L. Szafron. Monte Carlo simulation of strand passage in unknotted self-avoiding polygons. Master's thesis, University of Saskatchewan, 9 2000.

[106] Michael Lorne Szafron. *Knotting statistics after a local strand passage in unknotted self-avoiding polygons in* $\mathbb{Z}^3$. PhD thesis, University of Saskatchewan, 4 2009.

[107] M.L. Szafron and C.E. Soteros. The effect of juxtaposition angle on knot reduction in a lattice polygon model of strand passage. *Journal of Physics A: Mathematical and Theoretical*, 44(32):322001, 7 2011.

[108] M.L. Szafron and C.E. Soteros. Knotting probabilities after a local strand passage in unknotted self-avoiding polygons. *Journal of Physics A: Mathematical and Theoretical*, 44(24):245003, 5 2011.

[109] Myint Tin. Comparison of some ratio estimators. *Journal of the American Statistical Association*, 309(309):294–307, 3 1965.

[110] E.J. van Rensburg and A. Rechnitzer. On the universality of knot probability ratios. *Journal of Physics A: Mathematical and Theoretical*, 44(16):162002, 3 2011.

[111] E.J. Janse van Rensburg and A. Rechnitzer. Atmospheres of polygons and knotted polygons. *Journal of Physics A: Mathematical and Theoretical*, 41(10):105002, 2 2008.

[112] E.J. Janse van Rensburg and A. Rechnitzer. Generalized atmospheric sampling of knotted polygons. *Journal of Knot Theory and Its Ramifications*, 20(8):1145–1171, 6 2011.

[113] E.J. Janse van Rensburg and S.G. Whittington. The BFACF algorithm and knotted polygons. *Journal of Physics A: Mathematical and General*, 24(23):5553–5567, 12 1991.

[114] Carlo Vanderzande. *Lattice Models of Polymers*. Cambridge University Press, Cambridge, United Kingdom, 1998.

[115] Mariel Vazquez, Sean D. Colloms, and De Witt Sumners. Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single three-dimensional topological pathway. *Journal of Molecular Biology*, 346(2):493–504, 2005.

[116] Alexander V. Vologodskii. Circular DNA. In J. Anthony Semlyen, editor, *Cyclic Polymers*, chapter 2, pages 47–83. Springer, Netherlands, 2002. https://www.biophysics.org/Portals/1/PDFs/Education/Vologodskii.pdf.

[117] J.D. Watson and F.H.C. Crick. A structure for deoxyribose nucleic acids. *Nature*, 421(6921):397–398, 4 1953.

[118] Makoto Yamashita. Reidemeister move 1. `https://commons.wikimedia.org/wiki/File:Reidemeister_move_1.png`, 9 2006. CC BY-SA 3.0.

[119] Makoto Yamashita. Reidemeister move 2. `https://commons.wikimedia.org/wiki/File:Reidemeister_move_2.png`, 9 2006. CC BY-SA 3.0.

[120] Makoto Yamashita. Reidemeister move 3. `https://commons.wikimedia.org/wiki/File:Reidemeister_move_3.png`, 9 2006. CC By-SA 3.0.

[121] Jie Yan, Marcelo O. Magnasco, and John F. Marko. A kinetic proofreading mechanism for distanglement of DNA by topoisomerases. *Nature*, 401(6756):932–935, 10 1999.

Appendix A

Appendix: Analysis

A.1 $N_{max}(K)$ Graphs

Graphs of $N_{max}(K)$ graphs for starting knot types $3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+$, and 5_2^-. For the the determined $N_{max}(K)$ values see Table 4.1 in Section 4.1.1.

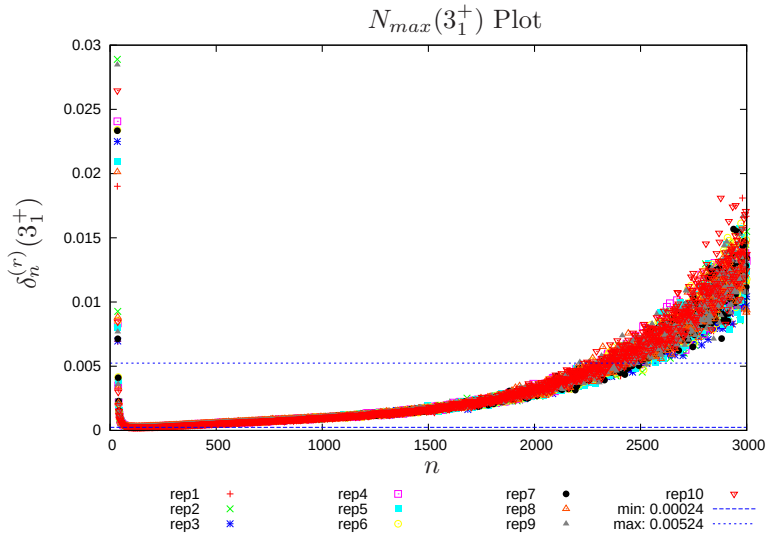


Figure A.1: Plot of the relative standard error starting with 3_1^+ to determine $N_{max}(3_1^+)$.

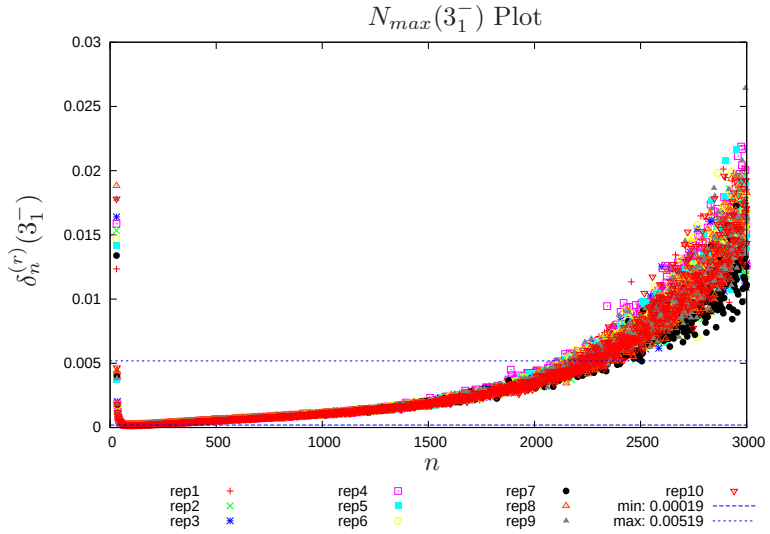


Figure A.2: Plot of the relative standard error starting with 3_1^- to determine $N_{max}(3_1^-)$.

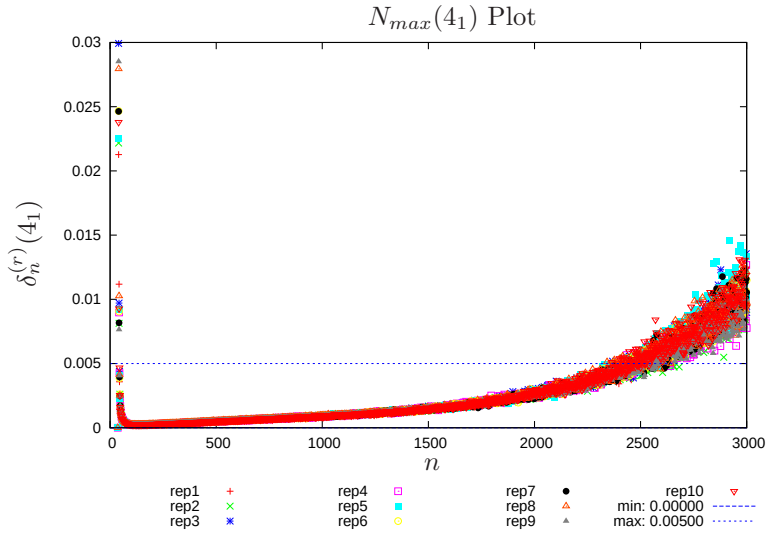


Figure A.3: Plot of the relative standard error starting with 4_1 to determine $N_{max}(4_1)$.

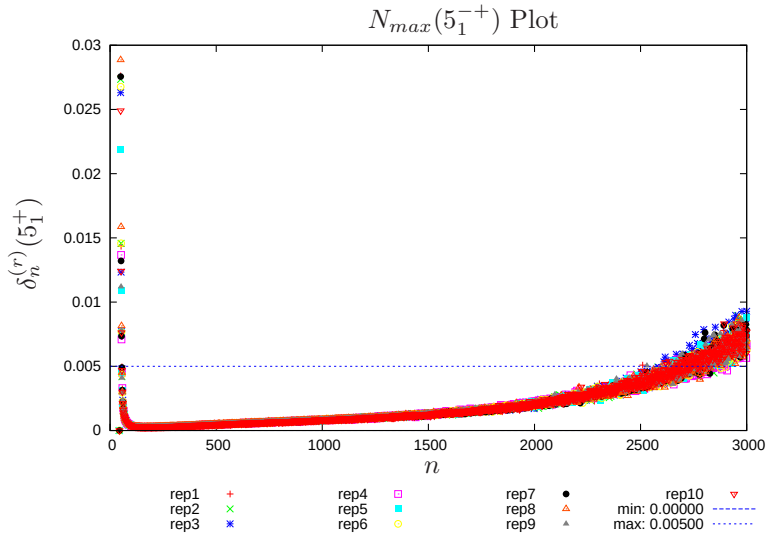


Figure A.4: Plot of the relative standard error starting with 5_1^+ to determine $N_{max}(5_1^+)$.

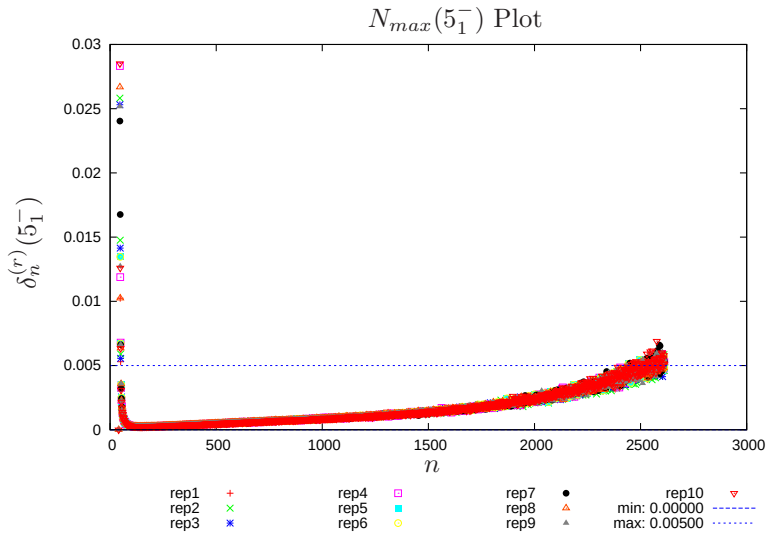


Figure A.5: Plot of the relative standard error starting with 5_1^- to determine $N_{max}(5_1^-)$.

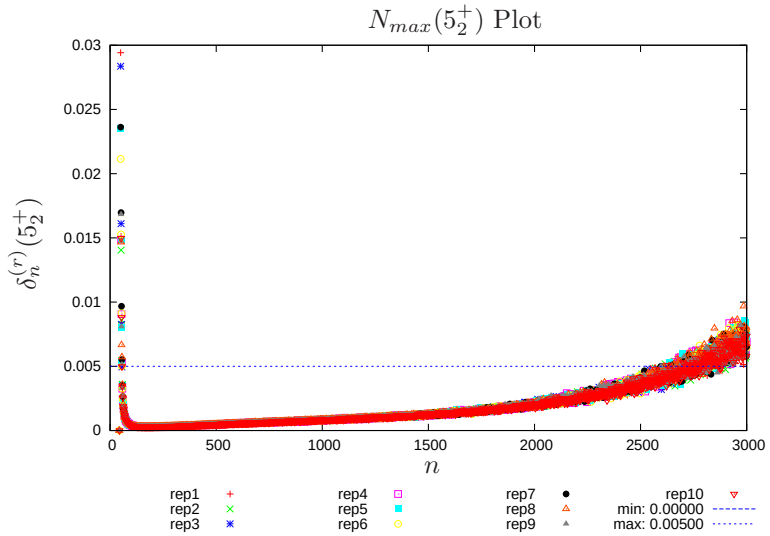


Figure A.6: Plot of the relative standard error starting with 5_2^+ to determine $N_{max}(5_2^+)$.

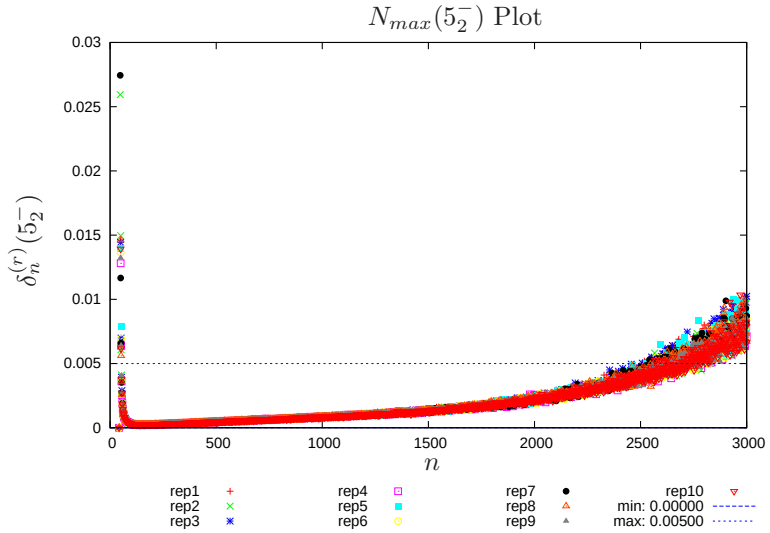


Figure A.7: Plot of the relative standard error starting with 5_2^- to determine $N_{max}(5_2^-)$.

A.2 Warm-up Analysis

Graphs of warm-up analysis for 5_1^- replications two to ten. For the determine k^*, k_*, and k values see Table 4.2 in Section 4.1.2.

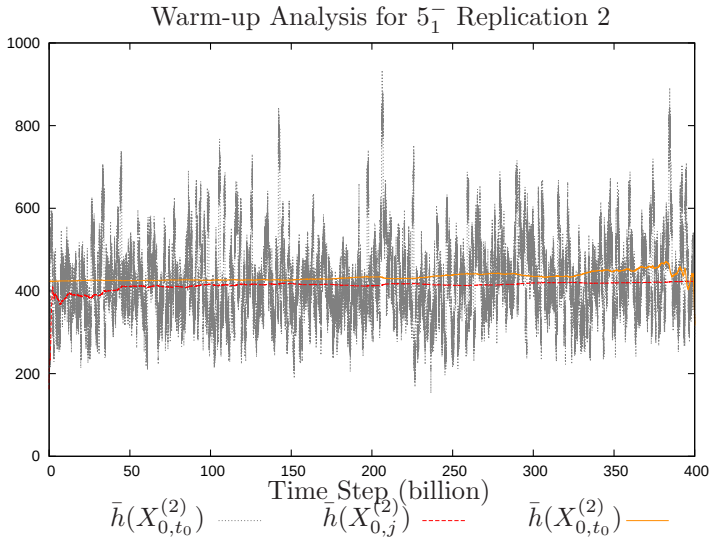


Figure A.8: Plot of warm-up analysis for replication two. $\left(h(X_j^{(2)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(2)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{2()}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

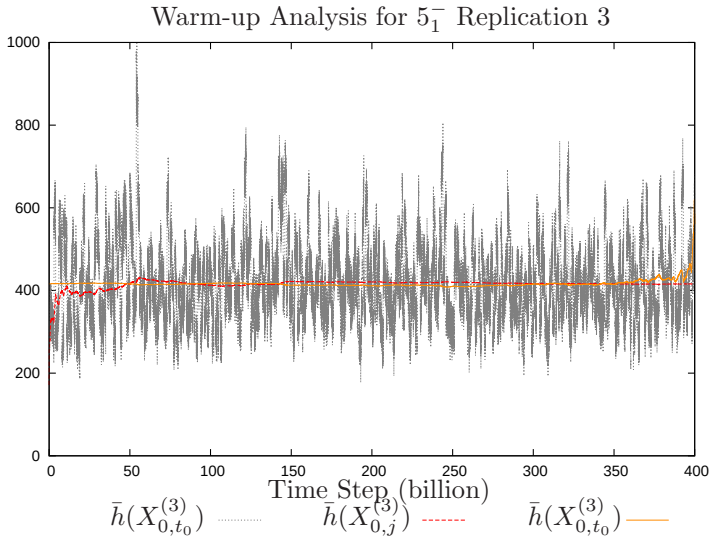


Figure A.9: Plot of warm-up analysis for replication three. $\left(h(X_j^{(3)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(3)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(3)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

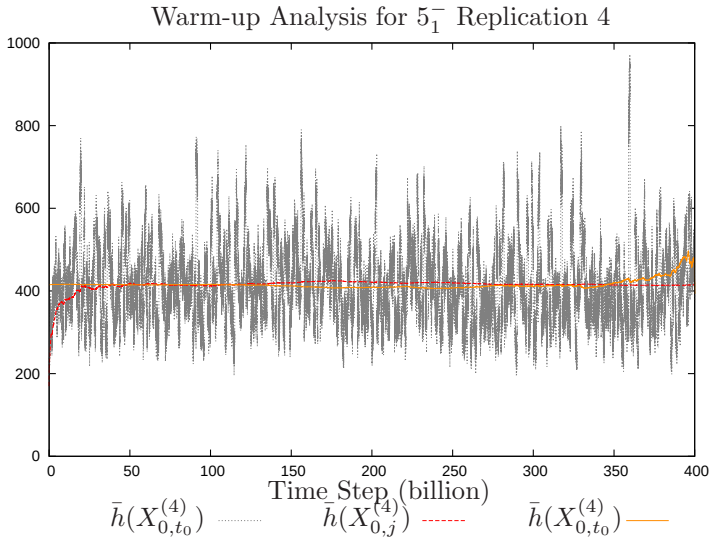


Figure A.10: Plot of warm-up analysis for replication four. $\left(h(X_j^{(4)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(4)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(4)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

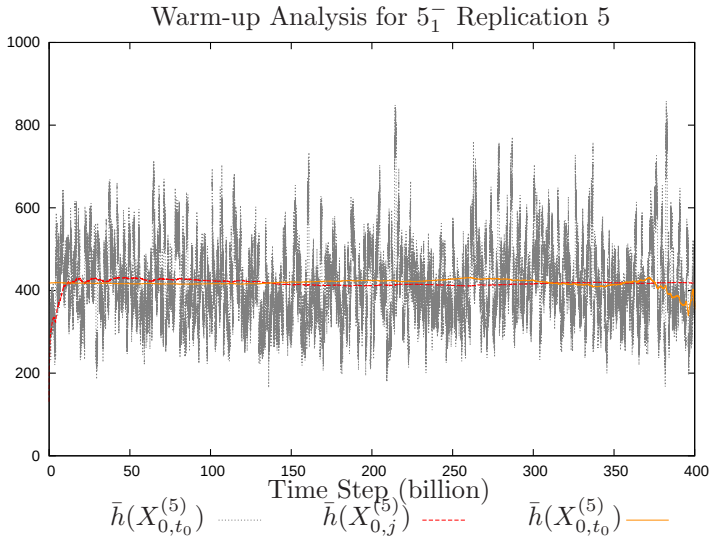


Figure A.11: Plot of warm-up analysis for replication five. $\left(h(X_j^{(5)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(5)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(5)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

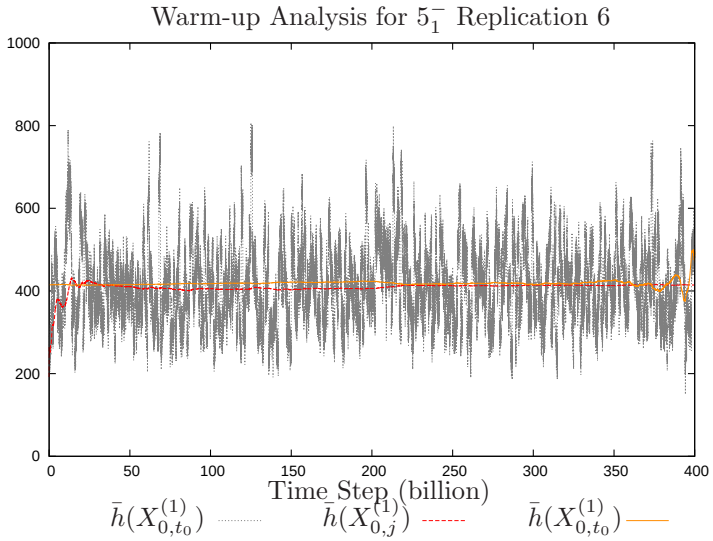


Figure A.12: Plot of warm-up analysis for replication six. $\left(h(X_j^{(6)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(6)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(6)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

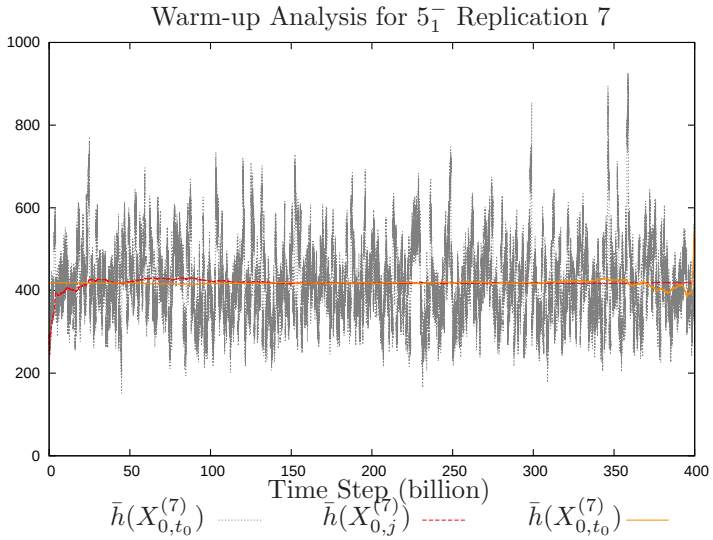


Figure A.13: Plot of warm-up analysis for replication seven. $\left(h(X_j^{(7)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(7)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(7)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

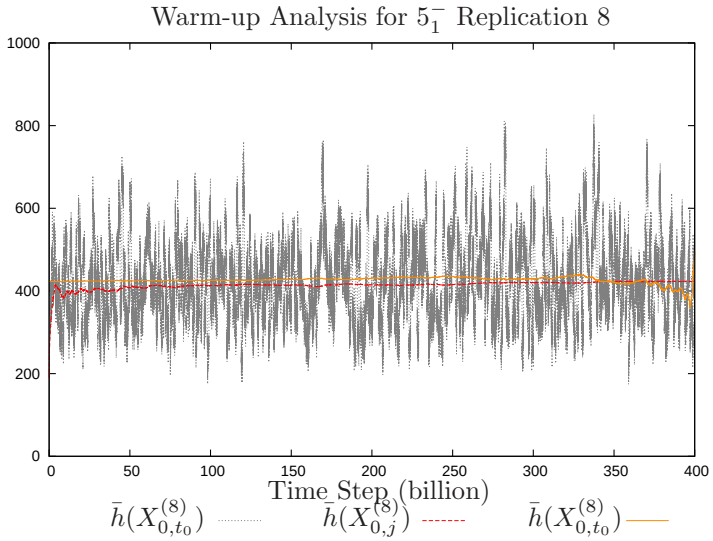


Figure A.14: Plot of warm-up analysis for replication eight. $\left(h(X_j^{(8)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(8)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(8)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

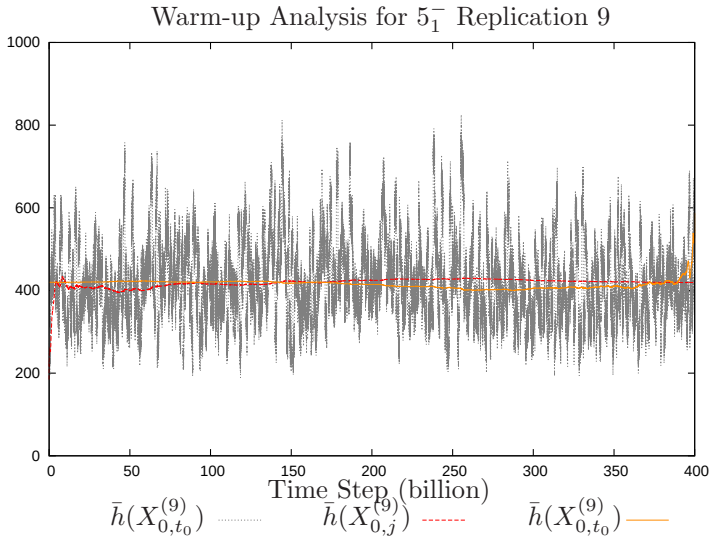


Figure A.15: Plot of warm-up analysis for replication nine. $\left(h(X_j^{(9)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(9)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(9)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

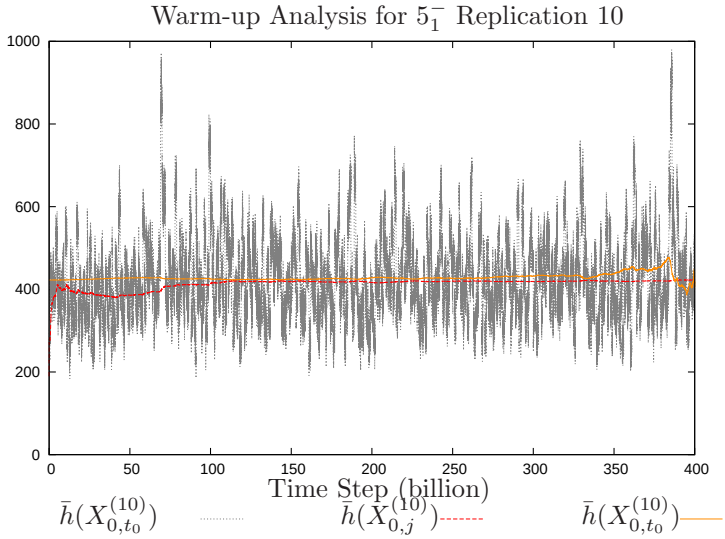


Figure A.16: Plot of warm-up analysis for replication ten. $\left(h(X_j^{(10)}), j = 0, \ldots, t_0\right)$ is the average polygon length over all 14 chains, $\left(\bar{h}(X_{0,j}^{(10)}), j = 0, \ldots, t_0\right)$ is the first j column average, and $\left(\bar{h}(X_{j,t_0}^{(10)}), j = 0, \ldots, t_0\right)$ is the the last j column average. Each is plotted over all 400 billion time steps.

Appendix B

Appendix: Link Transitions

B.1 Link Transitions

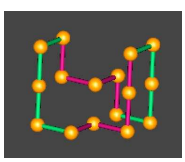
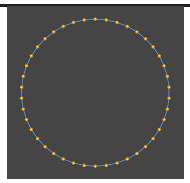
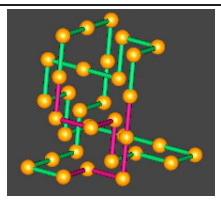
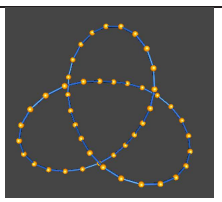
The following tables contain all observed transition probabilities given a successful strand exchange occurs. The first column (SE) of each table contains the observed knots given an IKE occurs or observed links given a DLE occurs. If an IKE occurs, column two (K/L) contains the corresponding link of the DLE. If a DLE occurs, column two (K/L) contains the corresponding knot of the IKE. Column three (L/K) is the link or knot type for the opposite move of that shown in column one for the example polygon in column six *i.e.* if a IKE occurs column three is the link resulting from a DLE on a example smallest polygon. Column four (Size) is the smallest polygon observed from the given strand exchange. Column five (Block Size)) is the grouped-n block size estimate for the given transition probability. As in Section 3.5.6 for *Case 2* knots from *Case 1*, Γ_{SE,C_1}, must be removed. Grouped-n block sizes were also calculated when Γ_{SE,C_1} knots had been removed but the same block size determined. Block sizes can only be determined when there is sufficient data for the transition probability. When insufficient data occurred, column four will have '-". Column six is a lattice picture (Lattice Picture) of one smallest size for the given strand exchange. Column seven (Representative) is a representative of the given knot or link type. It is a lattice polygon that has been reduced in size and relaxed in order to better see the shape of the knot or link. The orientation the representative knot and link does not necessarily correspond to the given knot of link type displayed with the positive or negative or letter from a to d. Representative polygons are only given for knot and link types were grouped-n blocking has been done.

As stated in Section 2.3 a link that is linked together is display with $l(K \cup K')$, while a knot that is unlinked is displayed with $u(K \cup K')$. There are two other knots not discussed earlier, $n(K \cup K')$ and $o(K \cup K')$. These knot are a knot linked with ϕ where there are four crossing between K and ϕ due to a twisting of one of the knots. 'comp.' and 'l_comp.' represent knots and links which contain the original knot, *i.e.* knots that are not Γ_{SE,C_1}. 'unk.' and 'l_unk.' represent knots and links that were not able to be identified.

B.2 Knot Type ϕ Observed Transitions

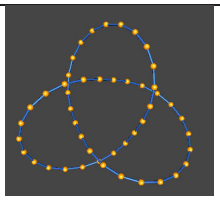
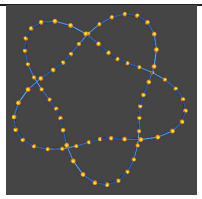
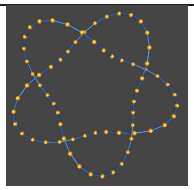
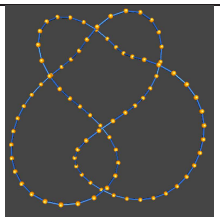
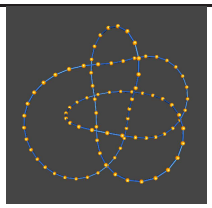
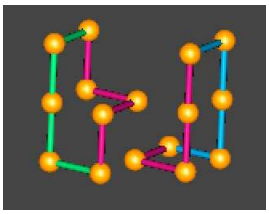
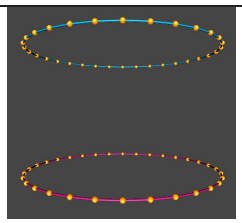
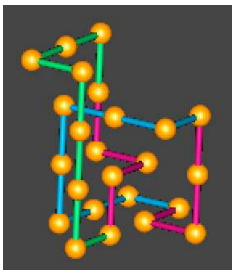
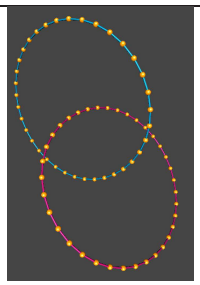
[Knot Type Unknot Observed Transitions]

Table B.1: Table showing all observed knot and link types after a strand exchange starting with ϕ.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	0_1^2	14	50		
IKE	3_1^+	$2_1^2 b$	32	80		

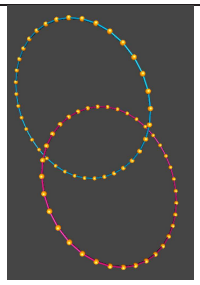
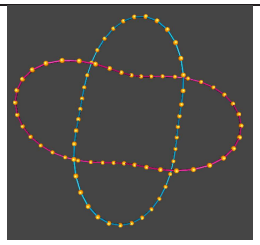
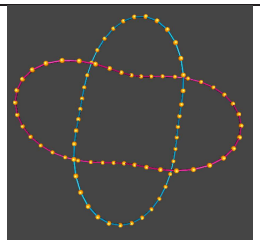
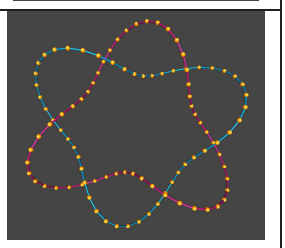
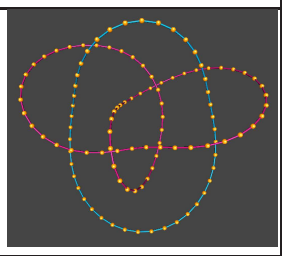
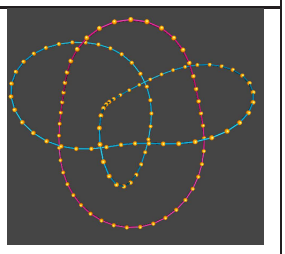
Continued on next page

Table B.1 – *ϕ knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
IKE	3_1^-	$4_1^2 d$	60	–	–	
IKE	5_1^+	$4_1^2 a$	138	–	–	
IKE	5_1^-	$7_7^2 b$	366	–	–	
IKE	5_2^-	$l3.1c$	226	–	–	
IKE	8_{20}^-	$7_8^2 c$	1,196	–	–	
DLE	0_1^2	0_1	14	50		
DLE	$2_1^2 a$	0_1	22	50		

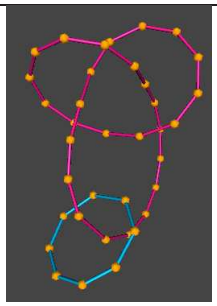
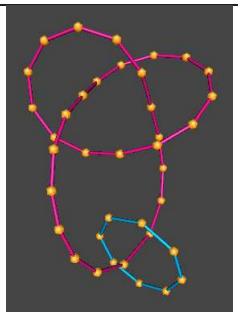
Continued on next page

Table B.1 – *ϕ knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$2_1^2 b$	3_1^+	32	–	–	
DLE	$4_1^2 a$	5_1^+	138	–	–	
DLE	$4_1^2 d$	3_1^-	60	–	–	
DLE	$6_1^2 c$	5_1^-	452	–	–	
DLE	$7_7^2 b$	5_1^-	366	–	–	
DLE	$7_8^2 c$	8_{20}^-	1,196	–	–	

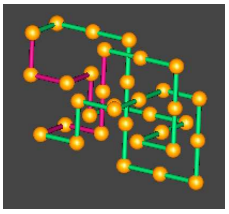
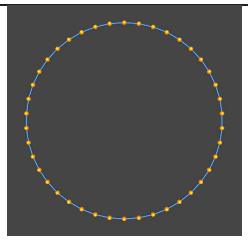
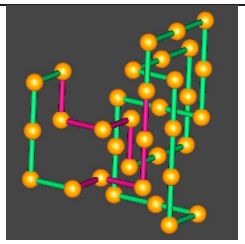
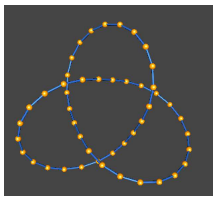
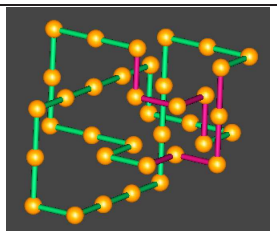
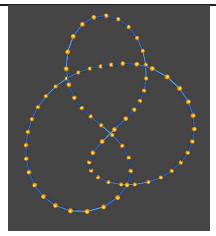
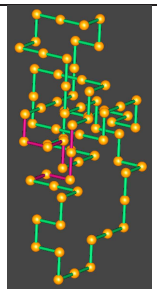
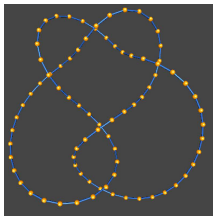
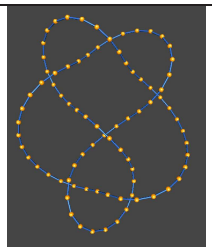
Continued on next page

Table B.1 – *ϕ knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(3_1^+ \cup \phi)$	5_1^+	656	–	–	
DLE	$l(3_1^- \cup \phi)$	5_2^-	226	–	–	
DLE	l_unk.	0_1	668	-	-	-
-	s	-	-	50	-	-

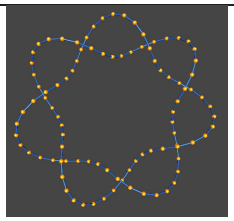
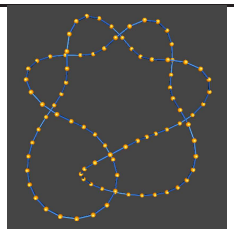
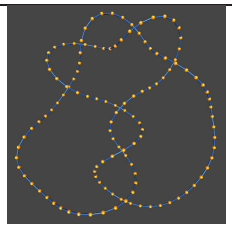
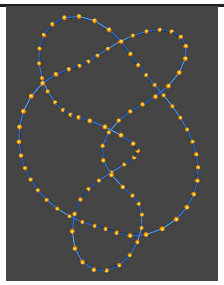
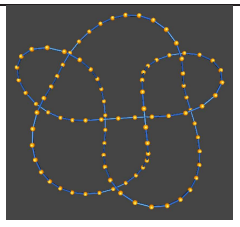
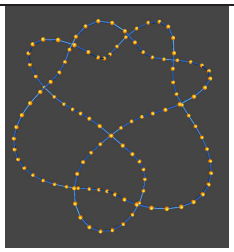
B.3 Knot Type 3_1^+ Observed Transitions

Table B.2: Table showing all observed knot and link types after a strand exchange starting with 3_1^+.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	$4_1^2 b$	32	60		
IKE	3_1^+	$u(3_1^+ \cup \phi)$	32	70		
IKE	4_1	$2_1^2 a$	34	60		
IKE	5_2^-	$6_2^2 a$	70	70		
IKE	6_2^+	$5_1^2 a$	74	–	–	

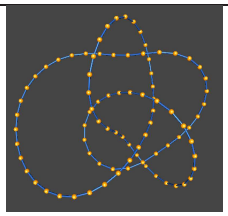
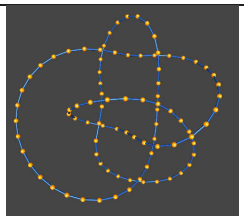
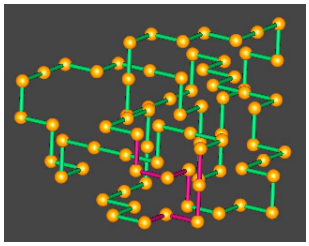
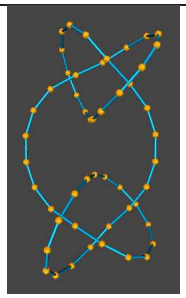
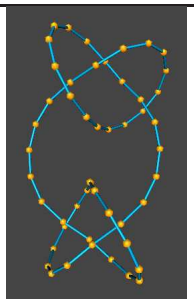
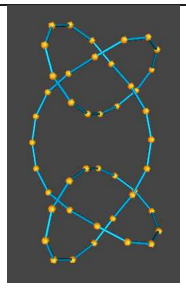
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_1^+	$l(5_1^+ \cup \phi)$	208	–	–	
IKE	7_2^+	$9_{47}^2 a$	670	–	–	
IKE	7_3^+	$9_{44}^2 a$	746	–	–	
IKE	7_5^+	$l(5_2^+ \cup \phi)$	356	–	–	
IKE	7_7^+	$9_{54}^2 b$	638	–	–	
IKE	8_2^+	$9_{53}^2 a$	714	–	–	

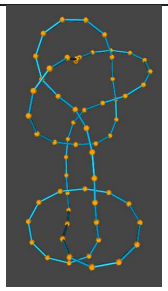
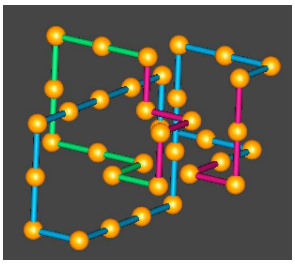
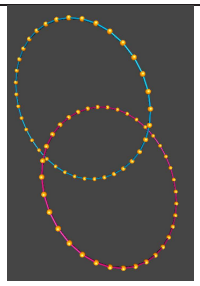
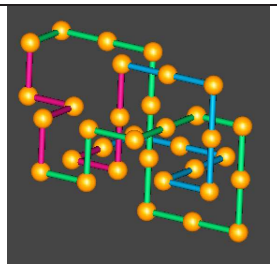
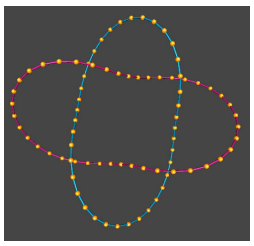
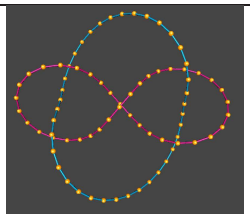
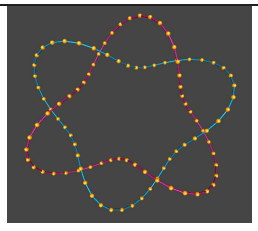
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{19}^-	$7_1^2 b$	218	–	–	
IKE	8_{21}^-	$6_3^2 d$	162	–	–	
IKE	$3_1^+ \# 3_1^+$	$l(3_1^+ \cup \phi)$	74	80		
IKE	$3_1^- \# 3_1^-$	$6_1^2 c$	82	–	–	
IKE	$3_1^+ \# 3_1^-$	$n(3_1^+ \cup \phi)$	122	–	–	

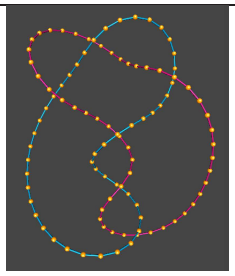
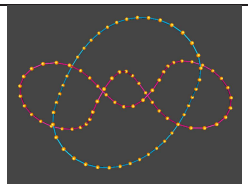
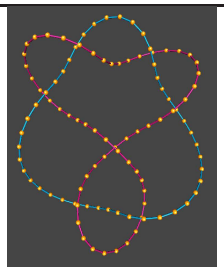
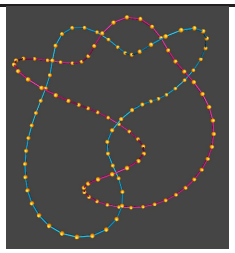
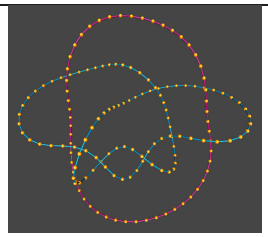
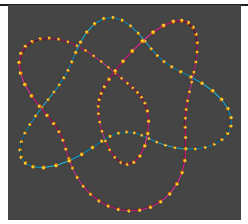
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	$3_1^+\#5_1^+$	$l(3_1^+\#3_1^+ \cup\phi)$	296	–	–	
IKE	Γ_{IKE,C_1}	-	-	60	-	-
IKE	comp.	-	-	80	-	-
DLE	2_1^2a	4_1	34	60		
DLE	4_1^2b	0_1	34	60		
DLE	5_1^2a	6_2^+	74	–	–	
DLE	6_1^2c	$3_1^-\#3_1^-$	82	–	–	

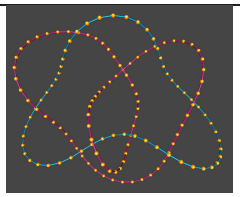
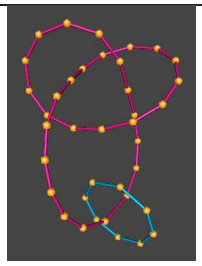
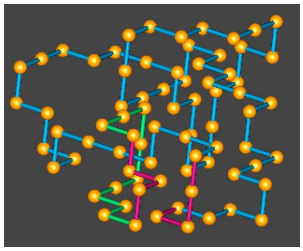
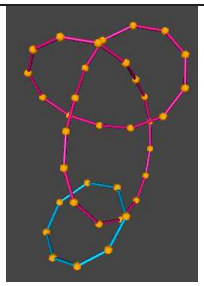
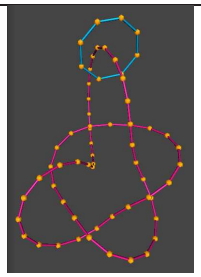
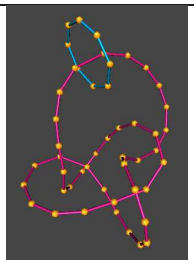
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	6_2^2a	5_2^-	70	–	–	
DLE	6_3^2d	8_{21}^-	162	–	–	
DLE	7_1^2b	8_2^+	714	–	–	
DLE	8_2^2b	7_3^-	1,174	–	–	
DLE	9_{44}^2a	7_3^+	746	–	–	
DLE	9_{53}^2a	8_{19}^-	218	–	–	

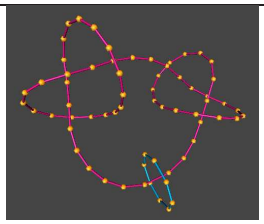
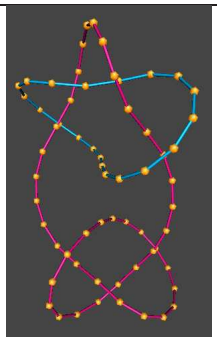
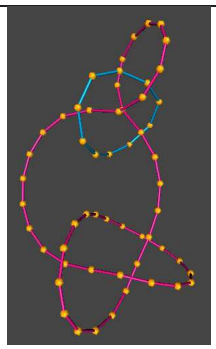
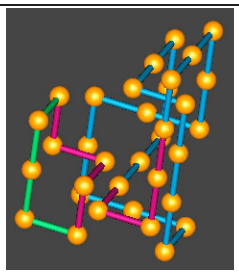
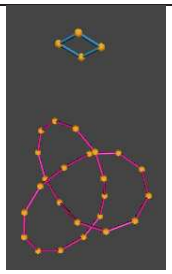
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{54}^2 b$	7_7^+	638	–	–	
DLE	$l(3_1^+ \cup \phi)$	3_1^+	44	–	–	
DLE	$l(3_1^+ \cup \phi)$	$3_1^+ \# 3_1^+$	74	80		
DLE	$l(5_1^+ \cup \phi)$	7_1^+	208	–	–	
DLE	$l(5_2^+ \cup \phi)$	7_5^+	356	–	–	

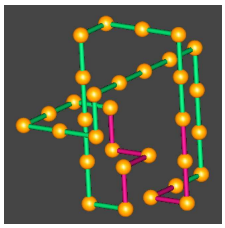
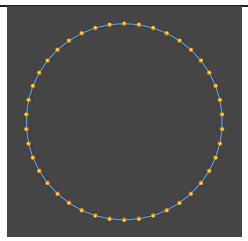
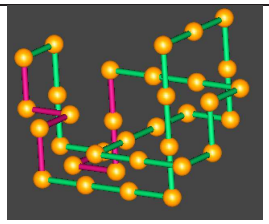
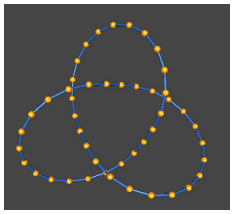
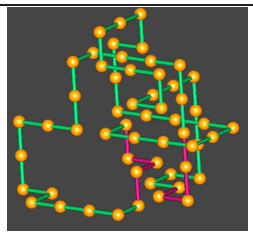
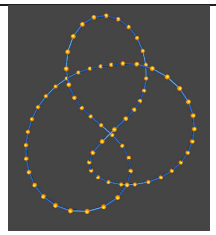
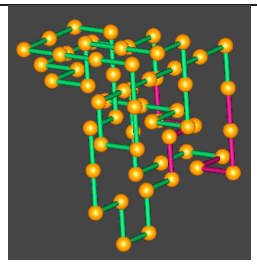
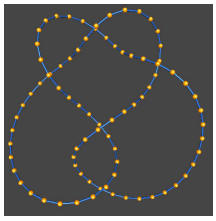
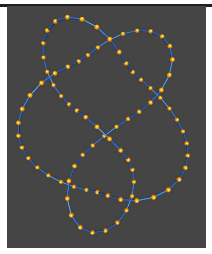
Continued on next page

Table B.2 – 3_1^+ *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(3_1^+\#3_1^+ \cup\phi)$	$3_1^+\#5_1^+$	296	–	–	
DLE	$n(3_1^+ \cup\phi)$	$3_1^+\#5_1^+$	300	–	–	
DLE	$n(3_1^+ \cup\phi)$	$3_1^+\#3_1^-$	122	–	–	
DLE	$u(3_1^+ \cup\phi)$	3_1^+	32	70		
DLE	l_unk.	3_1	184	-	-	-
DLE	Γ_{DLE,C_1}	-	-	60	-	-
DLE	l_comp.	-	-	70	-	-
-	s	-	-	70	-	-

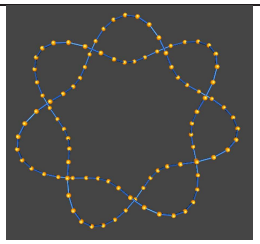
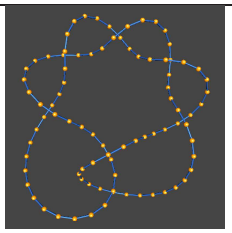
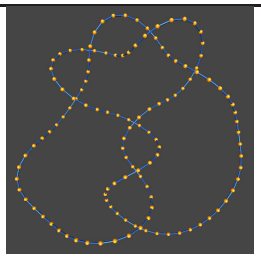
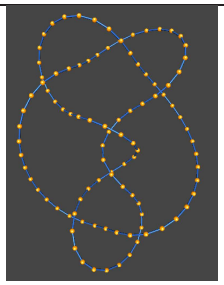
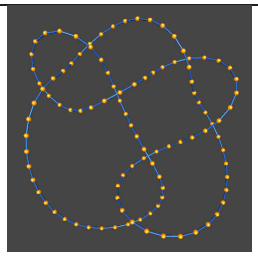
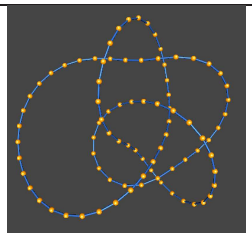
B.4 Knot Type 3_1^- Observed Transitions

Table B.3: Table showing all observed knot and link types after a strand exchange starting with 3_1^-.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	$2_1^2 a$	30	60		
IKE	3_1^+	$u(3_1^+ \cup \phi)$	32	70		
IKE	4_1	$5_1^2 a$	54	80		
IKE	5_2^-	$4_1^2 b$	52	70		
IKE	6_2^+	$7_1^2 b$	190	–	–	

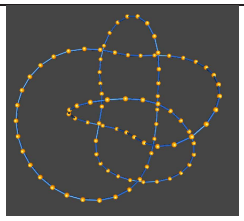
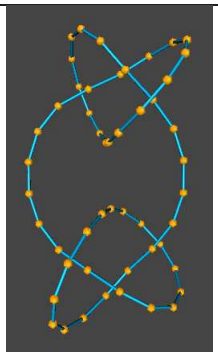
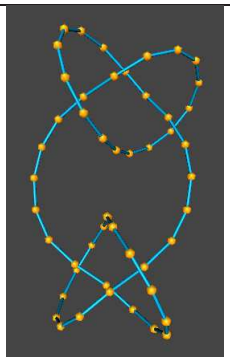
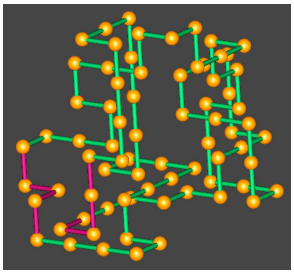
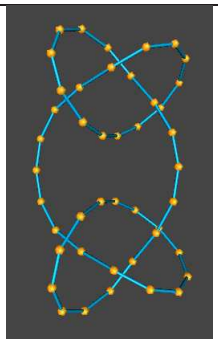
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_1^+	$9_{43}^2 d$	318	–	–	
IKE	7_2^+	$l(5_2^+ \cup \phi)$	208	–	–	
IKE	7_3^+	$l(5_1^+ \cup \phi)$	454	–	–	
IKE	7_5^+	$9_4^2 8c$	3234	–	–	
IKE	7_6^+	$7_4^2 c$	350	–	–	
IKE	8_{19}^-	$6_1^2 c$	164	–	–	

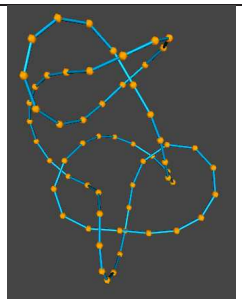
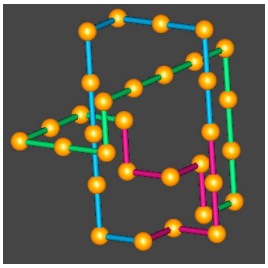
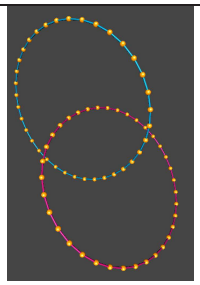
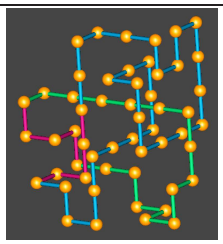
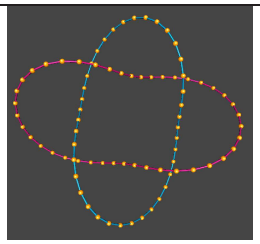
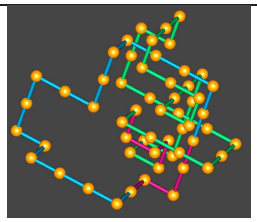
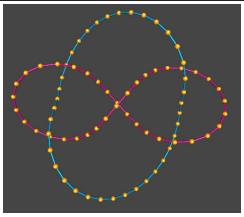
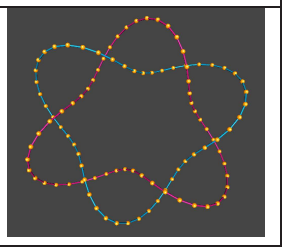
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{21}^-	$9_{54}^2 d$	1,796	–	–	
IKE	$3_1^+ \# 3_1^+$	$n(3_1^+ \cup \phi)$	140	–	–	
IKE	$3_1^- \# 3_1^-$	$6_3^2 d$	130	–	–	
IKE	$3_1^+ \# 3_1^-$	$l(3_1^+ \cup \phi)$	72	80		

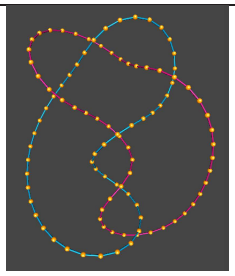
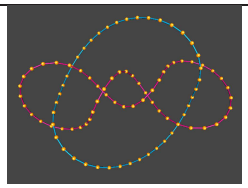
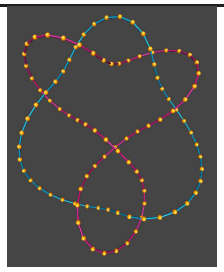
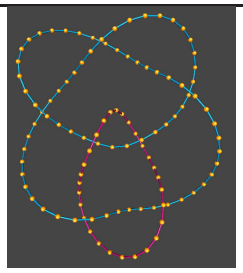
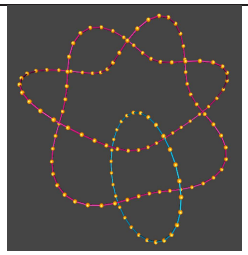
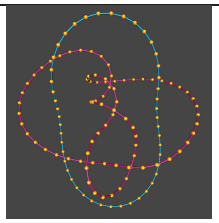
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	$3_1^+ \# 5_1^-$	$n(3_1^+ \cup \phi)$	172	–	–	
IKE	Γ_{IKE,C_1}	-	-	70	-	-
IKE	comp.	-	-	80	-	-
DLE	$2_1^2 a$	0_1	30	50		
DLE	$4_1^2 b$	5_2^-	52	80		
DLE	$5_1^2 a$	4_1	54	80		
DLE	$6_1^2 c$	8_{19}^-	164	–	–	

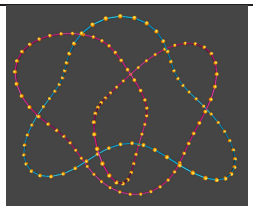
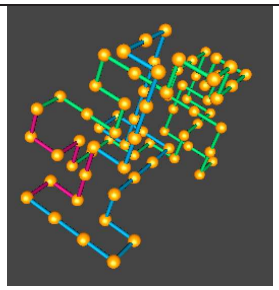
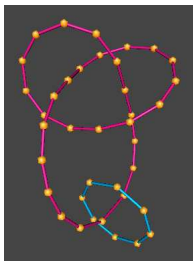
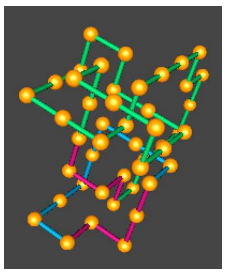
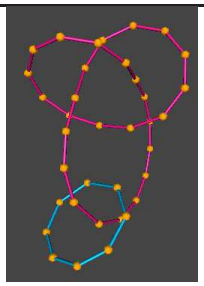
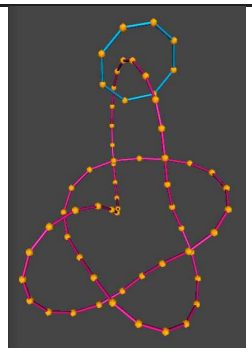
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	6_2^2a	7_3^-	164	–	–	
DLE	6_3^2d	$3_1^-\#3_1^-$	130	–	–	
DLE	7_1^2b	6_2^+	190	–	–	
DLE	7_4^2c	7_6^+	350	–	–	
DLE	9_{43}^2d	7_1^+	318	–	–	
DLE	9_{47}^2a	7_5^+	3,234	–	–	

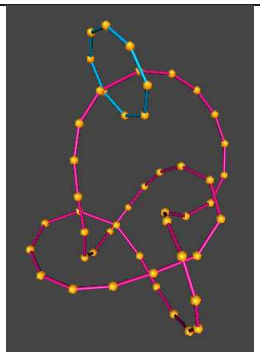
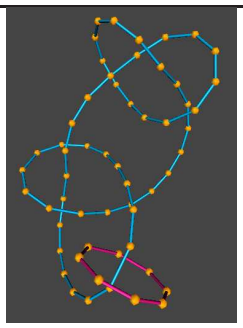
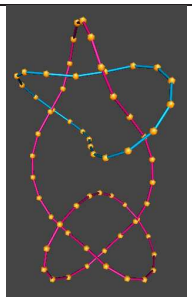
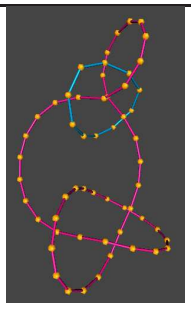
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{54}^2 d$	8_{21}^-	1,796	–	–	
DLE	$l(3_1^+ \cup \phi)$	$3_1^+ \# 3_1^-$	72	80		
DLE	$l(3_1^+ \cup \phi)$	3_1^+	44	80		
DLE	$l(5_1^+ \cup \phi)$	7_3^+	454	–	–	

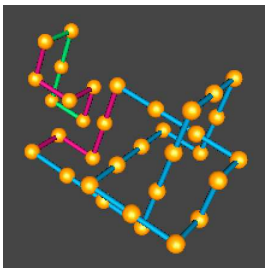
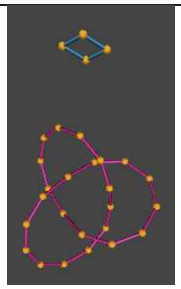
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$l(5_2^+ \cup \phi)$	7_2^+	208	–	–	
DLE	$l(3_1^+ \# 3_1^- \cup \phi)$	$3_1^+ \# 5_1^-$	1,388	–	–	
DLE	$n(3_1^+ \cup \phi)$	$3_1^+ \# 3_1^+$	140	–	–	
DLE	$n(3_1^+ \cup \phi)$	$3_1^+ \# 5_1^-$	172	–	–	

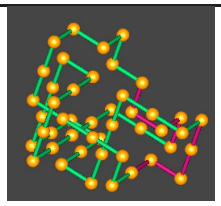
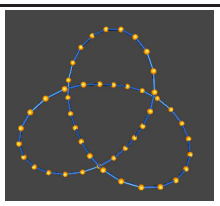
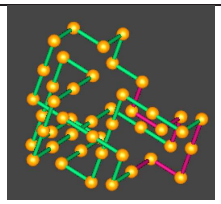
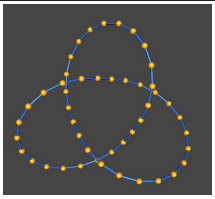
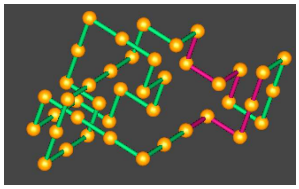
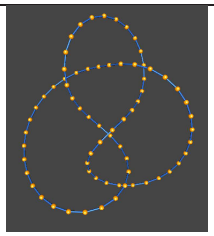
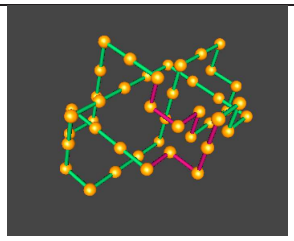
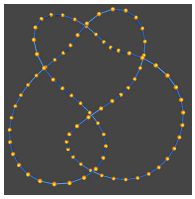
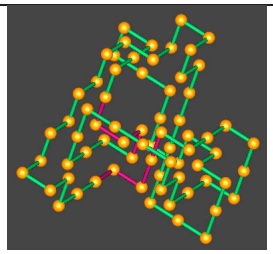
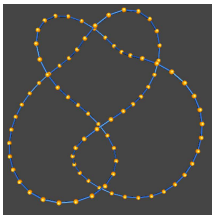
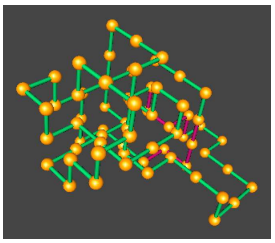
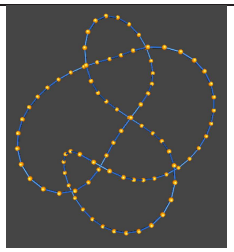
Continued on next page

Table B.3 – 3_1^- *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$u(3_1^+ \cup \phi)$	3_1^+	32	70		
DLE	l_unk.	0_1	182	–		
DLE	Γ_{DLE,C_1}	–	–	60	–	–
DLE	l_comp.	–	–	70	–	–
–	*s*	–	–	70	–	–

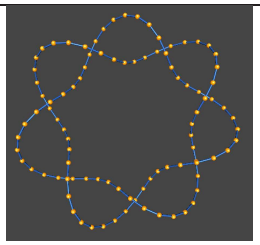
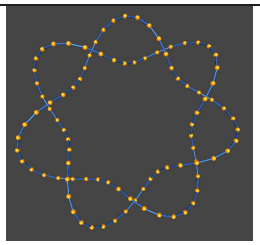
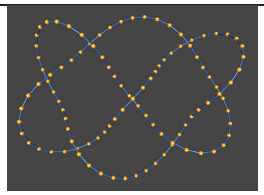
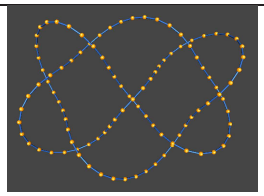
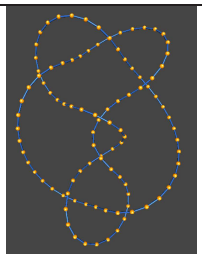
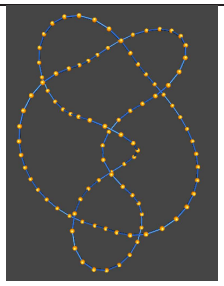
B.5 Knot Type 4_1 Observed Transitions

Table B.4: Table showing all observed knot and link types after a strand exchange starting with 4_1.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	3_1^+	5_1^2a	44	70		
IKE	3_1^-	2_1^2a	36	60		
IKE	4_1	$u(4_1 \cup \phi)$	40	80		
IKE	5_2^+	2_1^2b	42	70		
IKE	5_2^-	6_3^2d	68	80		
IKE	6_3^+	5_1^2c	66	80		

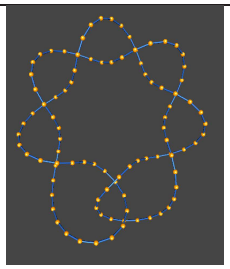
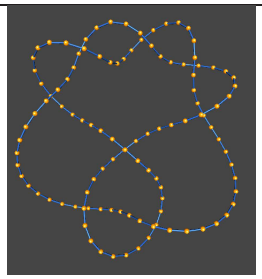
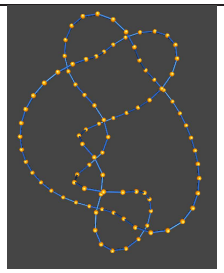
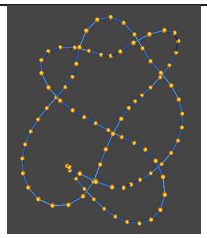
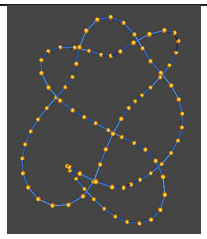
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_1^+	$n(3_1^+ \cup \phi)$	100	–	–	
IKE	7_1^-	$9_{49}^2 a$	180	–	–	
IKE	7_4^+	$9_{50}^2 a$	296	–	–	
IKE	7_4^-	$7_5^2 c$	196	–	–	
IKE	7_5^+	$9_{52}^2 a$	118	–	–	
IKE	7_5^-	$n(3_1^- \cup \phi)$	130	–	–	

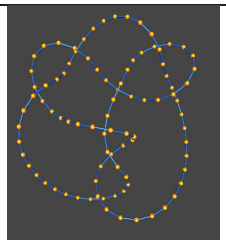
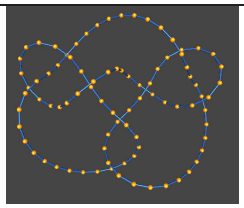
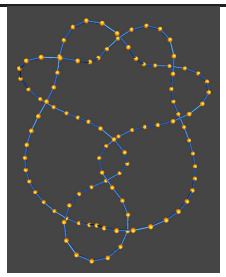
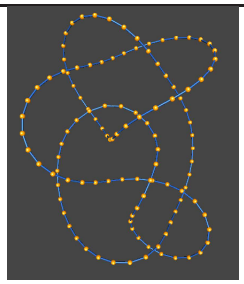
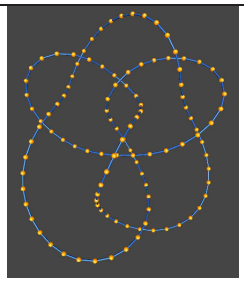
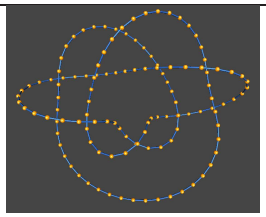
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_1^-	$l(6_1^- \cup \phi)$	196	–	–	
IKE	8_2^+	$l(6_2^+ \cup \phi)$	152	–	–	
IKE	8_6^+	$l(6_1^+ \cup \phi)$	292	–	–	
IKE	8_7^+	7_2^2c	218	–	–	
IKE	8_7^-	9_2^2a	394	–	–	

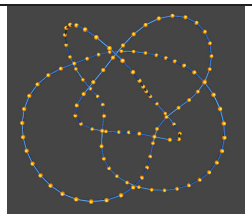
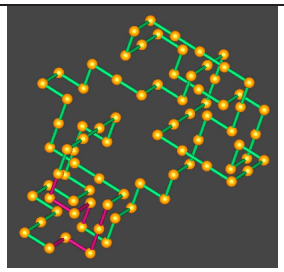
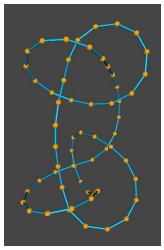
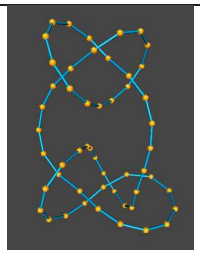
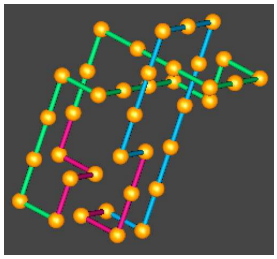
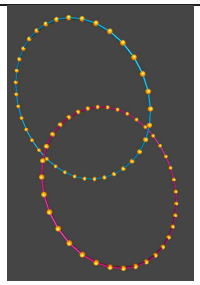
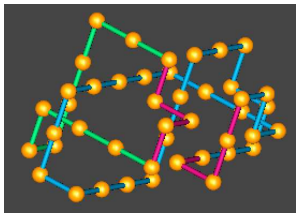
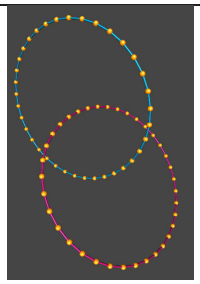
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{11}^{-}	$l(6_2^{-} \cup \phi)$	336	–	–	
IKE	8_{15}^{-}	$8_{11}^{2}b$	508	–	–	
IKE	9_{6}^{+}	$8_{3}^{2}b$	1,084	–	–	
IKE	9_{45}^{-}	$9_{58}^{2}a$	372	–	–	
IKE	9_{49}^{-}	$7_{5}^{2}b$	566	–	–	
IKE	10_{139}^{-}	$9_{49}^{2}d$	312	–	–	

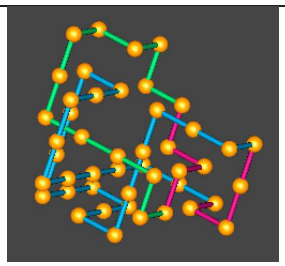
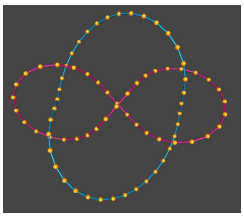
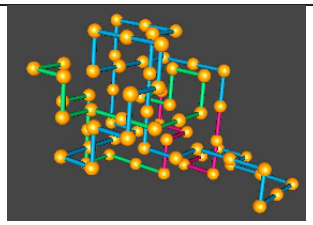
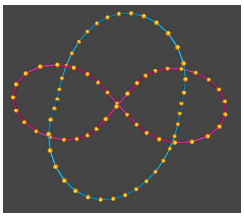
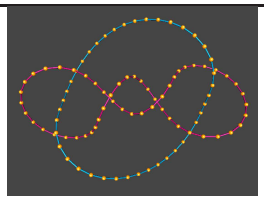
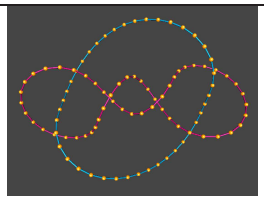
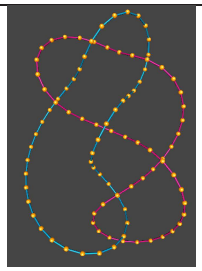
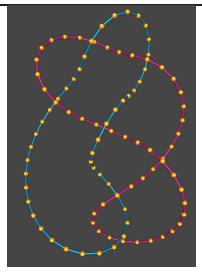
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10^-_{143}	$9^5_2 2d$	420	–	–	
IKE	unk.	$n(4_1 \cup \phi)$	238	–	–	
IKE	Γ_{IKE,C_1}	-	-	60	-	-
IKE	comp.	-	-	80	-	-
IKE	$3^+_1 \# 4_1$	$l(4_1 \cup \phi)$	80	80		
IKE	$3^-_1 \# 4_1$	$n(4_1 \cup \phi)$	216	–	–	
DLE	$2^2_1 a$	3^-_1	36	60		
DLE	$2^2_1 b$	5^+_2	42	70		

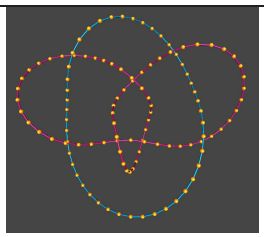
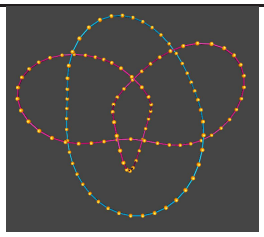
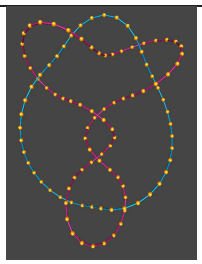
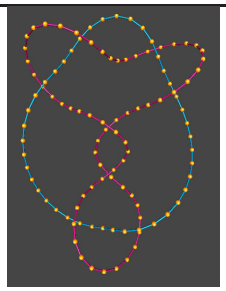
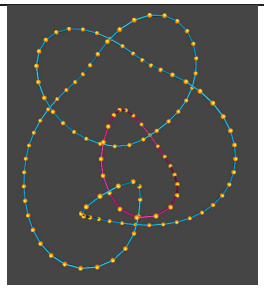
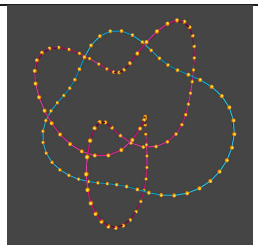
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	5_1^2a	3_1^+	44	70		
DLE	5_1^2c	6_3^+	66	80		
DLE	6_3^2a	7_5^+	120	–	–	
DLE	6_3^2d	5_2^-	68	–	–	
DLE	7_2^2b	6_3^+	130	–	–	
DLE	7_2^2c	8_7^+	218	–	–	

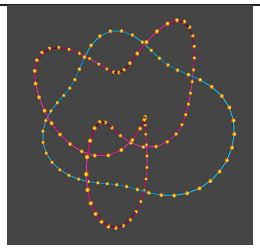
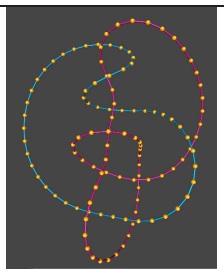
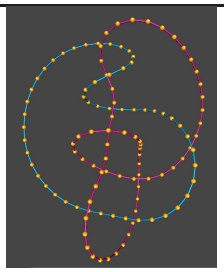
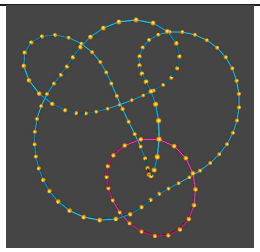
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$7_5^2 b$	9_{49}^-	566	–	–	
DLE	$7_5^2 c$	7_4^-	196	–	–	
DLE	$8_3^2 b$	9_6^+	1,084	–	–	
DLE	$8_3^2 c$	7_5^-	186	–	–	
DLE	$8_{11}^2 b$	8_{15}^-	508	–	–	
DLE	$9_{49}^2 a$	7_1^-	180	–	–	

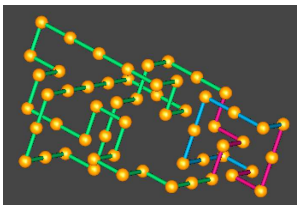
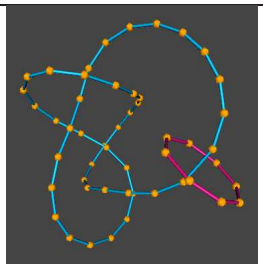
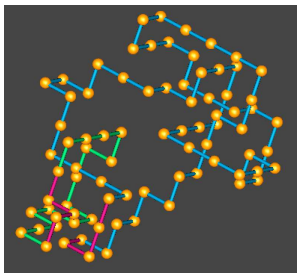
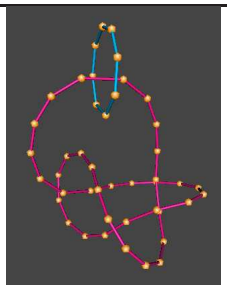
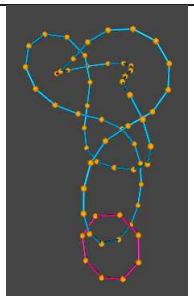
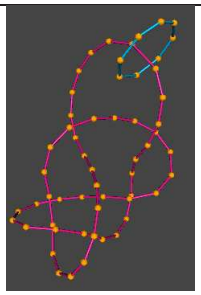
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9^2_{49}d$	10^-_{139}	312	–	–	
DLE	$9^2_{50}a$	7^+_4	296	–	–	
DLE	$9^2_{52}a$	7^+_5	118	–	–	
DLE	$9^2_{52}d$	10^-_{143}	420	–	–	
DLE	$9^2_{58}a$	9^-_{45}	372	–	–	

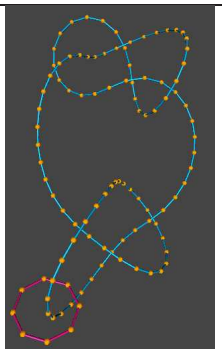
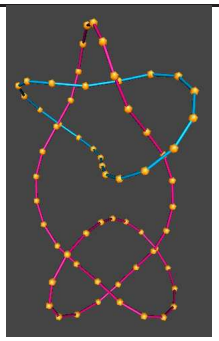
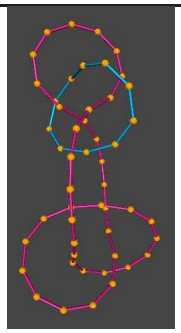
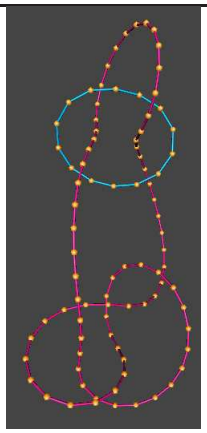
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(4_1 \cup \phi)$	4_1	54	80		
DLE	$l(4_1 \cup \phi)$	$3_1^+ \# 4_1$	80	80		
DLE	$l(6_1^+ \cup \phi)$	8_6^+	292	–	–	
DLE	$l(6_1^- \cup \phi)$	$8_1 s$	196	–	–	

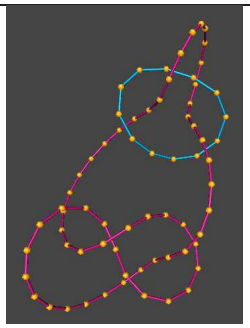
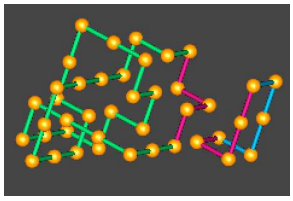
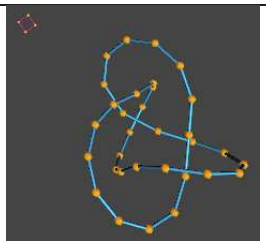
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$l(3_1^+ \# 4_1 \cup \phi)$	unk.	720	–	–	
DLE	$n(3_1^+ \cup \phi)$	7_1^+	100	–	–	
DLE	$n(3_1^- \cup \phi)$	7_5^-	130	–	–	
DLE	$n(4_1 \cup \phi)$	unk.	238	–	–	

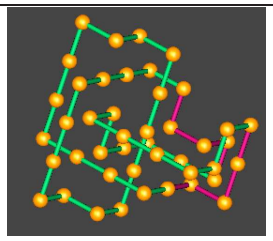
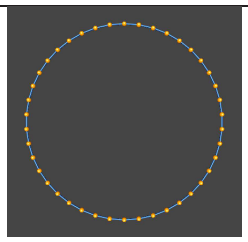
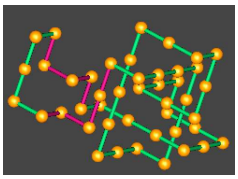
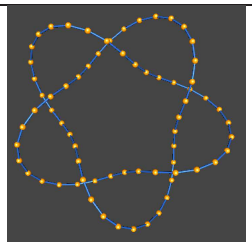
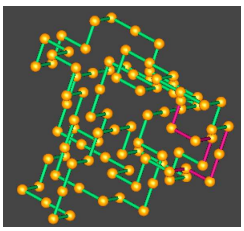
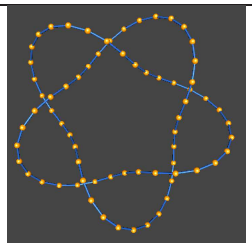
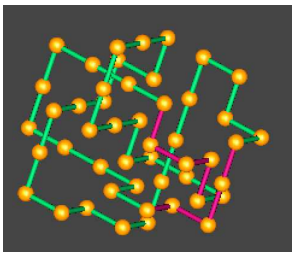
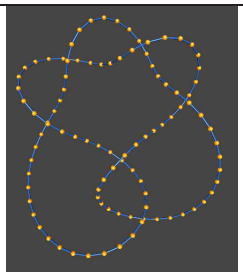
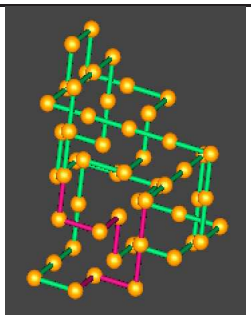
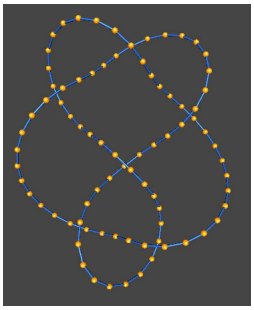
Continued on next page

Table B.4 – 4_1 *knot and link types after a strand exchange continued from previous page*

SE	$\boldsymbol{K/L}$	$\boldsymbol{L/K}$	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$n(4_1 \cup \phi$	$3_1^- \# 4_1$	176	–	–	
DLE	$u(4_1 \cup \phi)$	4_1	40	80		
DLE	l_unk.	3_1^-	188	–	–	–
DLE	Γ_{DLE,C_1}	–	–	60	–	–
DLE	l_comp.	–	–	80	–	–
–	s	–	–	70	–	–

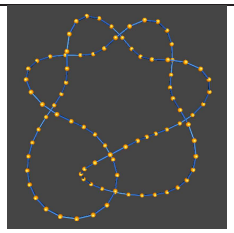
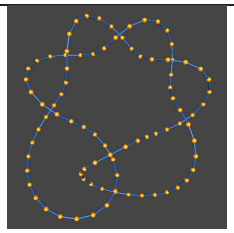
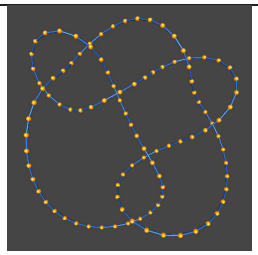
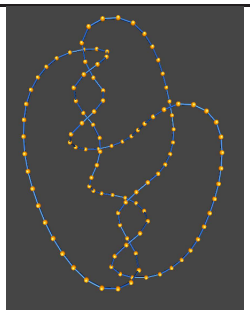
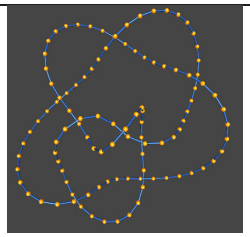
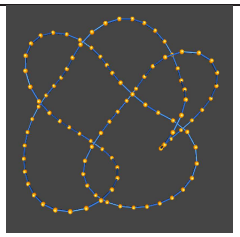
B.6 Knot Type 5_1^+ Observed Transitions

Table B.5: Table showing all observed knot and link types after a strand exchange starting with 5_1^+.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	$7_7^2 d$	42	70		
IKE	5_1^+	$u(5_1^+ \cup \phi)$	42	80		
IKE	5_1^-	$9_{61}^2 a$	90	100		
IKE	6_1^-	$4_1^2 b$	46	80		
IKE	6_2^+	$l(3_1^+ \cup \phi)$	50	80		

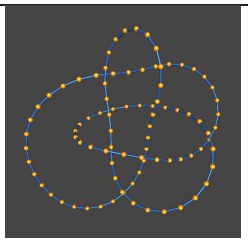
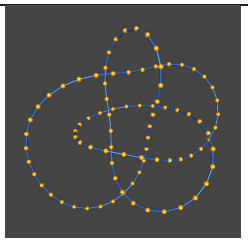
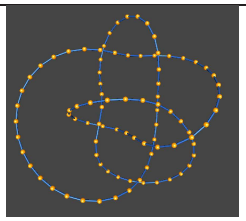
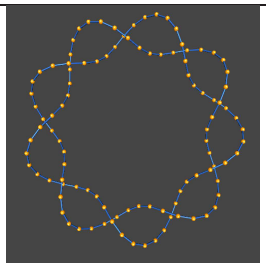
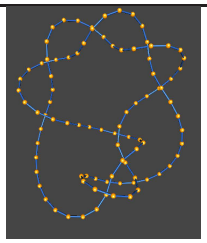
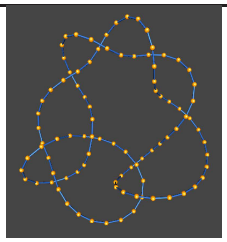
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_2^+	unk.	588	–	–	
IKE	7_2^-	$8_2^2 d$	108	–	–	
IKE	7_6^+	$l(5_2^+ \cup \phi)$	80	–	–	
IKE	8_4^+	$7_1^2 a$	112	–	–	
IKE	8_5^+	$7_4^2 c$	104	–	–	
IKE	8_{14}^-	$8_5^2 a$	284	–	–	

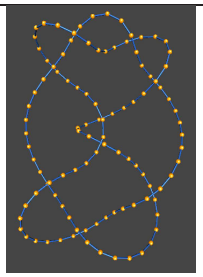
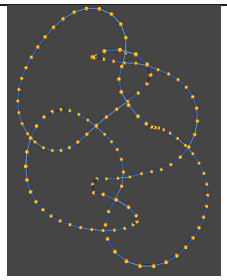
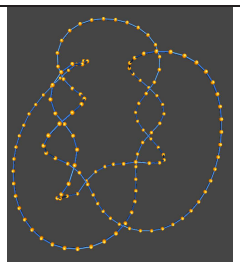
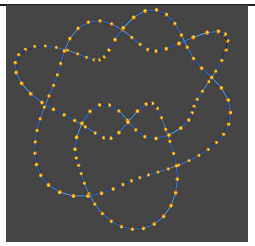
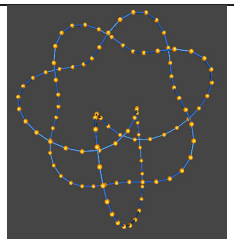
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{20}^+	$9_{60}^2 d$	94	–	–	
IKE	8_{20}^-	$9_{51}^2 c$	88	–	–	
IKE	8_{21}^-	$6_2^2 a$	84	–	–	
IKE	9_1^+	$l(7_1^+ \cup \phi)$	264	–	–	
IKE	9_4^-	unk.	1,926	–	–	
IKE	9_8^+	unk.	486	–	–	

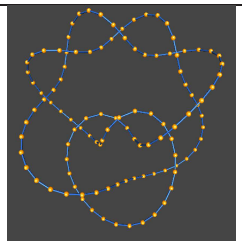
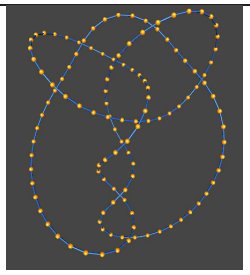
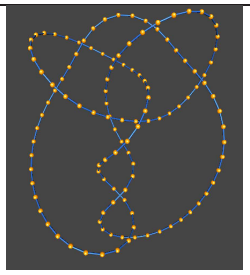
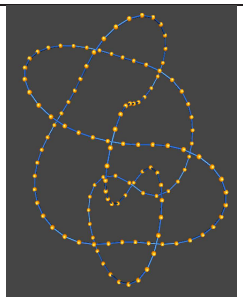
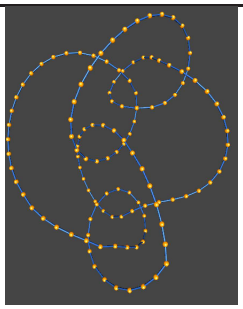
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	9_9^+	$l(7_3^- \cup \phi)$	310	–	–	
IKE	9_{16}^-	$l(7_5^+ \cup \phi)$	276	–	–	
IKE	10_8^+	$9_4^2 a$	476	–	–	
IKE	10_{46}^-	$9_{19}^2 c$	1,068	–	–	
IKE	10_{124}^-	unk.	198	–	–	

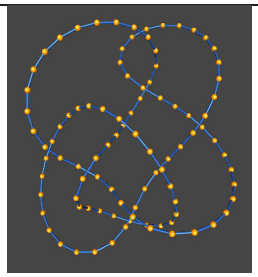
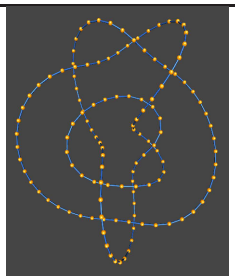
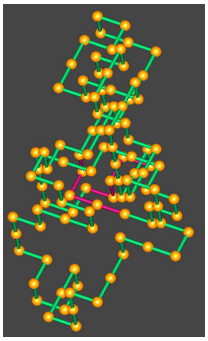
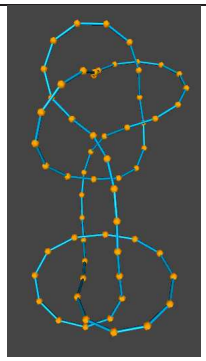
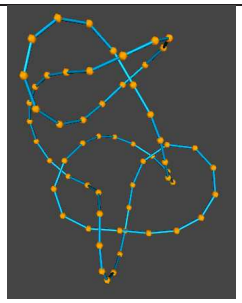
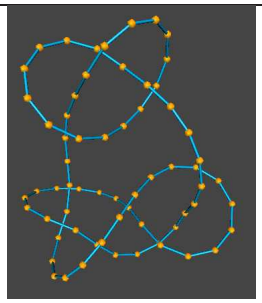
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_{126}^-	unk.	820	–	–	
IKE	10_{128}^+	unk.	564	–	–	
IKE	10_{128}^-	unk.	654	–	–	
IKE	10_{131}^-	8_5^2c	248	–	–	
IKE	10_{138}^-	8_7^2b	408	–	–	

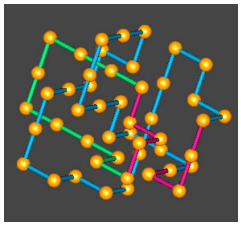
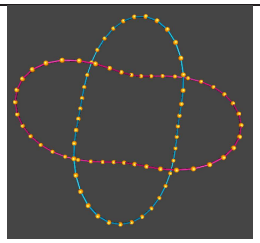
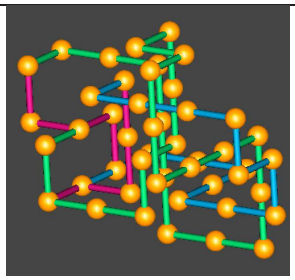
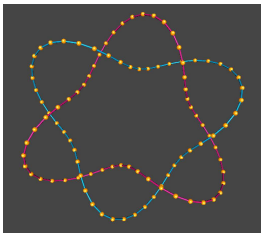
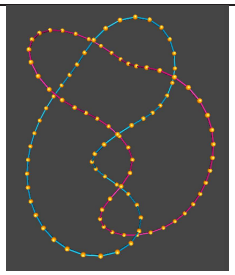
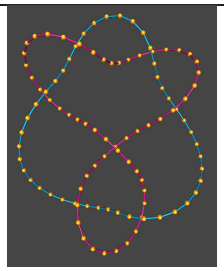
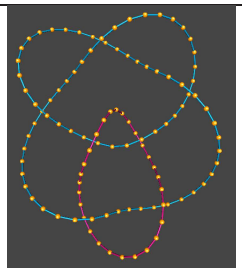
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_{150}^-	9_{56}^2c	206	–	–	
IKE	10_{152}^-	unk.	372	–	–	
IKE	unk.	9_{20}^2a	290	-	-	-
IKE	$3_1^+\#5_1^+$	$l(5_1^+ \cup \phi)$	102	100		
IKE	$3_1^-\#5_1^+$	unk.	138	–	–	
IKE	$3_1^-\#5_2^-$	8_2^2b	164	–	–	

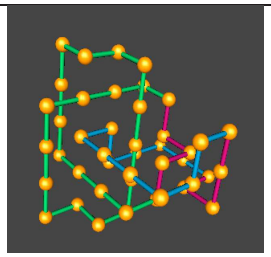
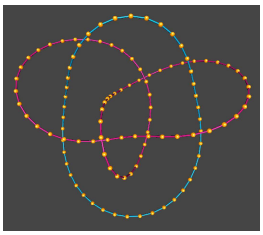
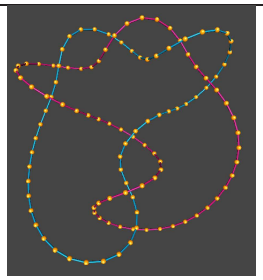
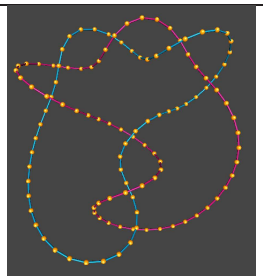
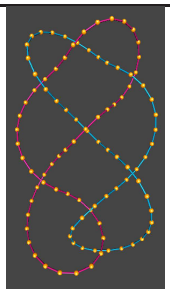
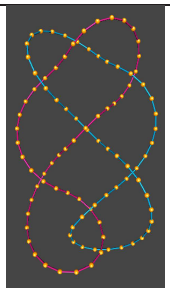
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
IKE	Γ_{IKE,C_1}	–	-	70	-	-
IKE	comp.	–	-	100	-	-
DLE	4_1^2b	6_1^-	46	70		
DLE	6_1^2a	0_1	44	70		
DLE	6_2^2a	8_{21}^-	84	–	–	
DLE	7_1^2a	8_4^+	112	–	–	
DLE	7_4^2c	8_5^+	104	–	–	

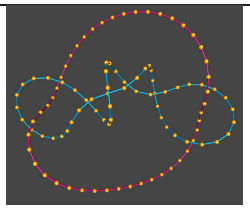
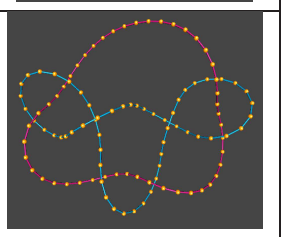
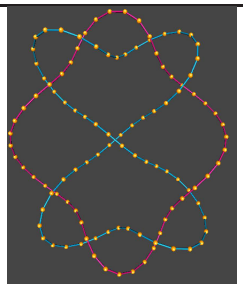
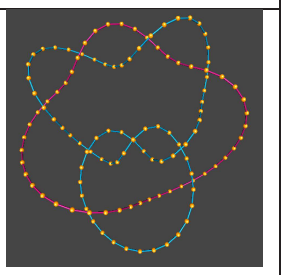
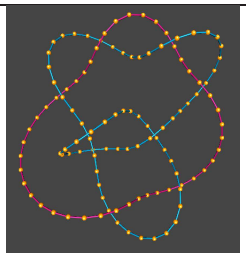
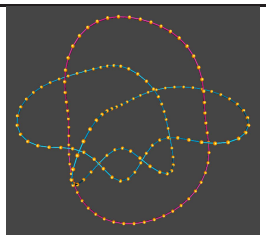
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$7_7^2 d$	0_1	42	70		
DLE	$8_2^2 b$	$3_1^- \# 5_2^-$	164	–	–	
DLE	$8_2^2 d$	7_2^-	108	–	–	
DLE	$8_5^2 a$	8_{14}^-	284	–	–	
DLE	$8_5^2 c$	10_{131}^-	248	–	–	

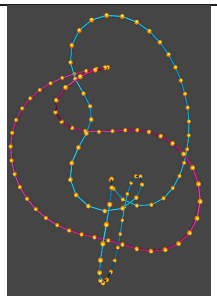
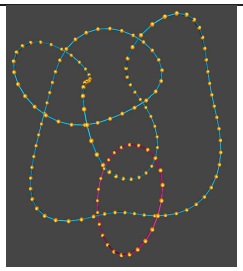
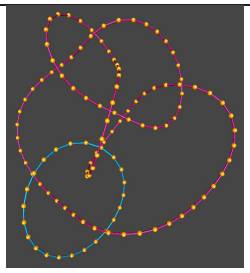
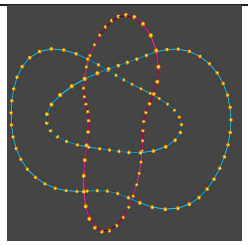
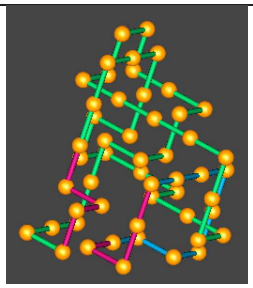
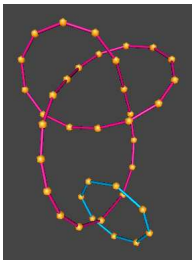
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	8_6^2a	$4_1\#4_1$	160	–	–	
DLE	8_7^2b	10_{138}^-	408	–	–	
DLE	9_4^2a	10_8^+	476	–	–	
DLE	9_{19}^2c	10_{46}^-	1,068	–	–	
DLE	9_{20}^2a	unk.	290	–	–	
DLE	9_{44}^2c	unk.	696	–	–	

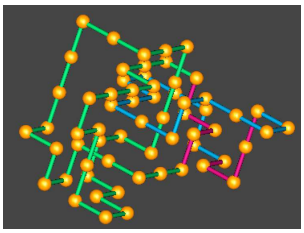
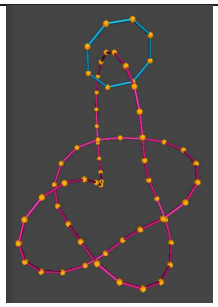
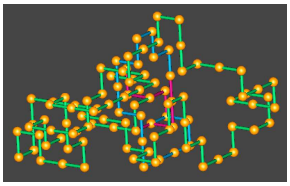
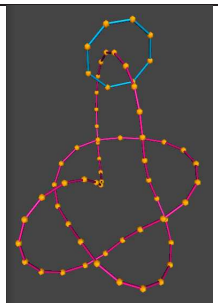
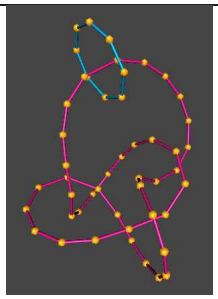
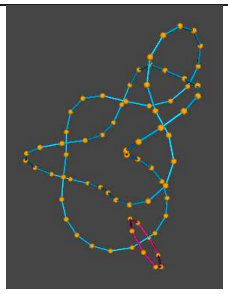
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{51}^2 c$	8_{20}^-	88	–	–	
DLE	$9_{56}^2 c$	10_{150}^-	206	–	–	
DLE	$9_{60}^2 d$	8_{20}^+	94	–	–	
DLE	$9_{61}^2 a$	5_1^-	90	–	–	
DLE	$l(3_1^+ \cup \phi)$	6_2^+	50	80		

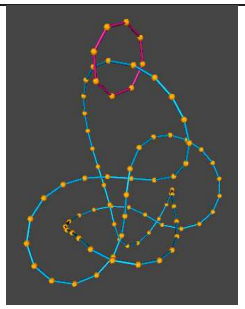
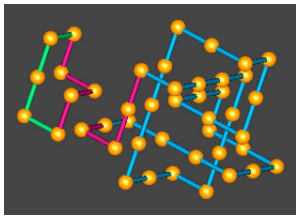
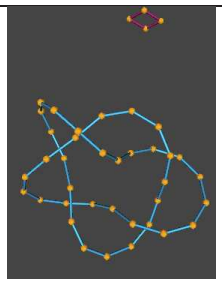
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	$\boldsymbol{K/L}$	$\boldsymbol{L/K}$	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(5_1^+ \cup \phi)$	5_1^+	62	100		
DLE	$l(5_1^+ \cup \phi)$	$3_1^+ \# 5_1^+$	102	100		
DLE	$l(5_2^+ \cup \phi)$	7_6^+	80	–	–	
DLE	$l(7_1^+ \cup \phi)$	9_1^+	264	–	–	

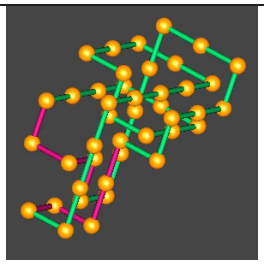
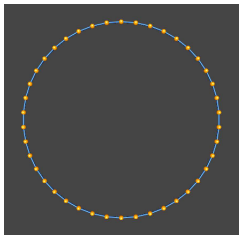
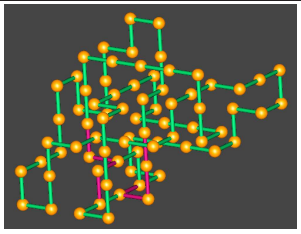
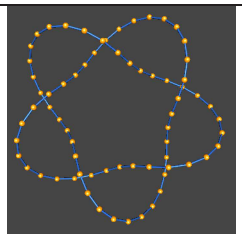
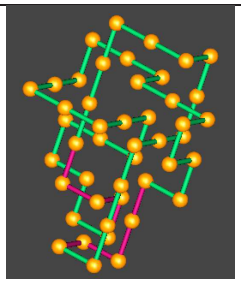
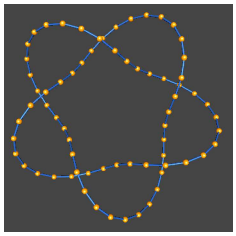
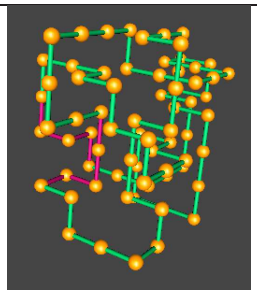
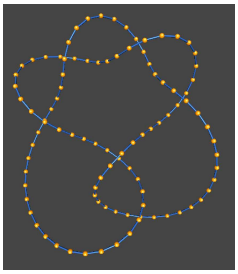
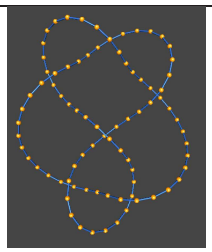
Continued on next page

Table B.5 – 5_1^+ *knot and link types after a strand exchange continued from previous page*

SE	$\boldsymbol{K/L}$	$\boldsymbol{L/K}$	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(7_3^- \cup \phi)$	9_9^+	310	–	–	
DLE	$l(7_5^+ \cup \phi)$	9_{16}^-	276	–	–	
DLE	$u(5_1^+ \cup \phi)$	5_1^+	42	80		
DLE	l_unk.	5_1^+	136	–	–	–
DLE	Γ_{DLE,C_1}	–	–	70	–	–
DLE	l_comp.	–	–	80	–	–
DLE	s	–	–	80	–	–

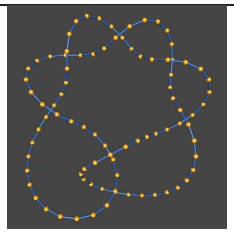
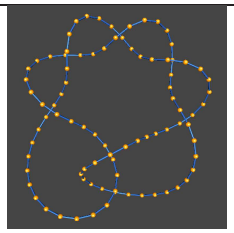
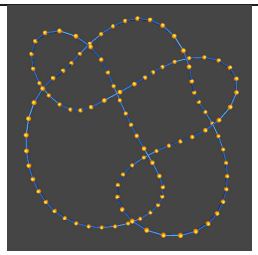
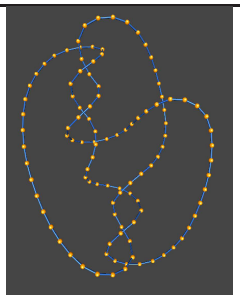
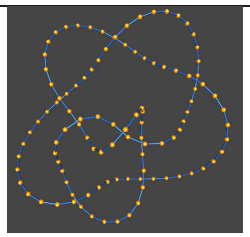
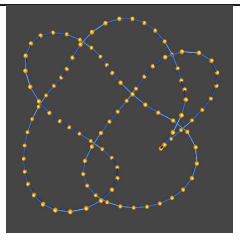
B.7 Knot Type 5_1^- Observed Transitions

Table B.6: Table showing all observed knot and link types after a strand exchange starting with 5_1^-.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	4_1^2c	40	70		
IKE	5_1^+	6_2^2b	72	80		
IKE	5_1^-	$u(5_1^- \cup \phi)$	46	100		
IKE	6_1^+	7_1^2d	70	100		
IKE	6_2^-	7_4^2a	74	–	–	

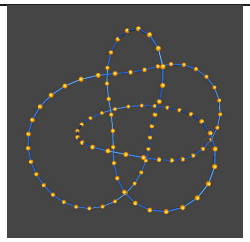
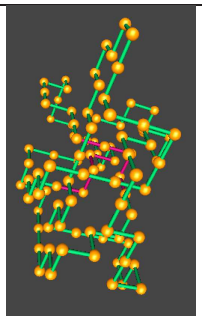
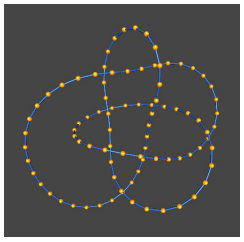
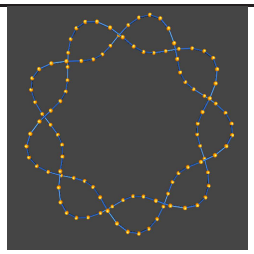
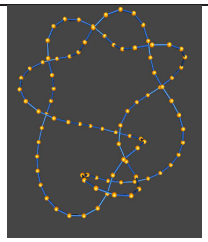
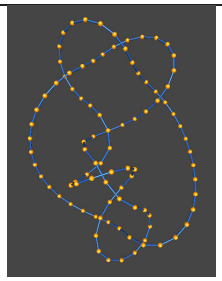
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_2^+	$6_1^2 d$	74	–	–	
IKE	7_2^-	$9_{44}^2 c$	248	–	–	
IKE	7_6^-	$9_{56}^2 a$	168	–	–	
IKE	8_4^-	$9_4^2 c$	292	–	–	
IKE	8_5^-	$9_{19}^2 b$	324	–	–	
IKE	8_{14}^+	unk.	254	–	–	

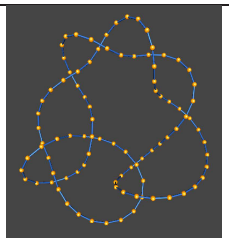
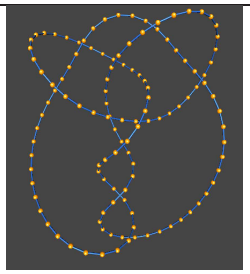
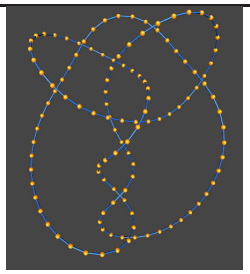
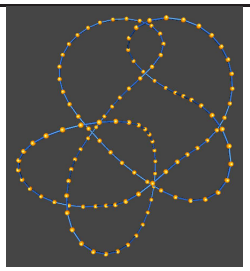
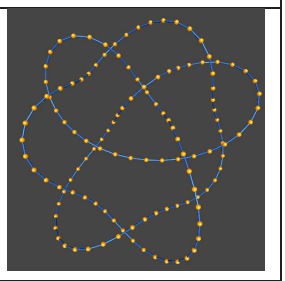
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{20}^+	7_7^2a	68	–	–	
IKE	8_{20}^-	$l(5_2^- \cup \phi)$	94	100		
IKE	9_1^-	unk.	386	–	–	
IKE	9_4^-	$l(7_3^+ \cup \phi)$	372	–	–	
IKE	9_7^-	$l(7_5^- \cup \phi)$	406	–	–	

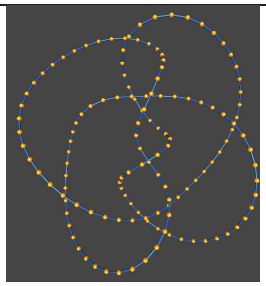
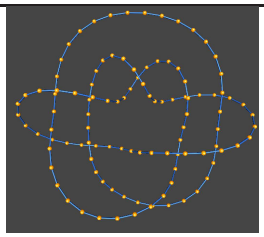
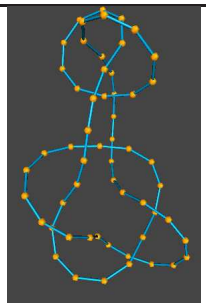
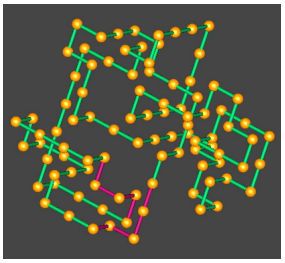
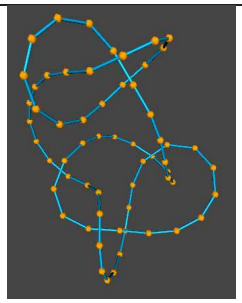
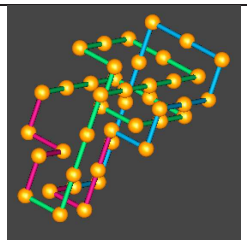
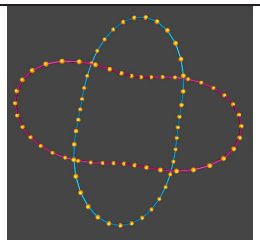
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	9_8^-	9_{19}^2a	690	–	–	
IKE	10_{128}^+	unk.	202	–	–	
IKE	10_{128}^-	8_2^2c	358	–	–	
IKE	10_{133}^-	unk.	294	–	–	
IKE	10_{155}^+	unk.	656	–	–	

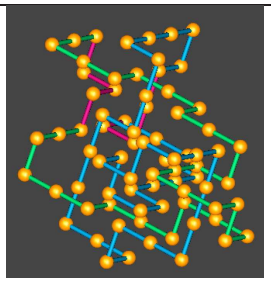
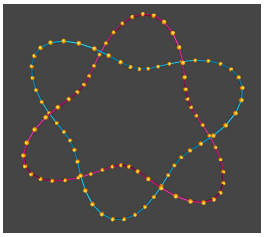
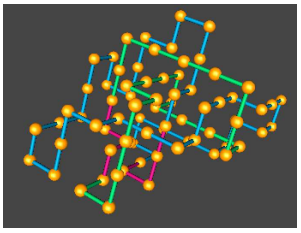
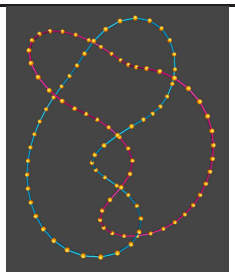
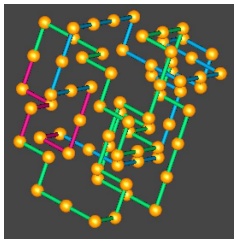
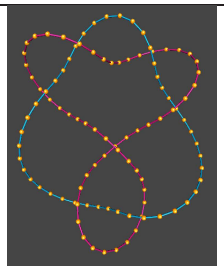
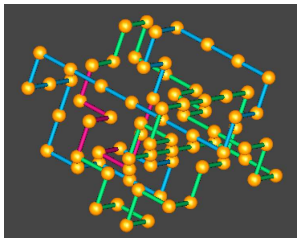
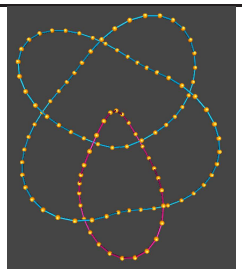
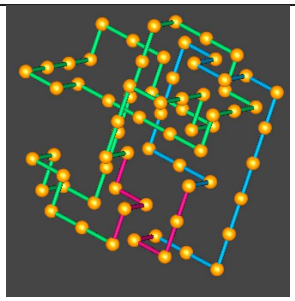
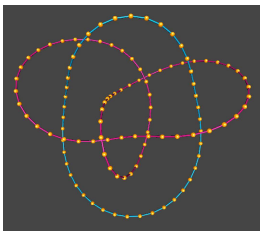
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_{160}^-	$8_5^2 d$	612	–	–	
IKE	10_{161}^+	$9_{61}^2 b$	178	–	–	
IKE	$3_1^+ \# 5_1^+$	unk.	140	–	–	
IKE	$3_1^+ \# 5_1^-$	$l(5_1^- \cup \phi)$	92	100		
IKE	unk.	unk.	260	–	–	–
IKE	Γ_{IKE,C_1}	–	–	70	–	–
IKE	comp.	–	–	100	–	–
DLE	$4_1^2 c$	0_1	40	70		

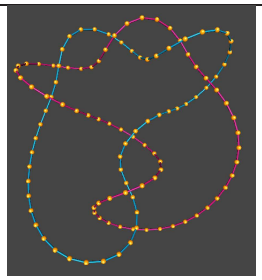
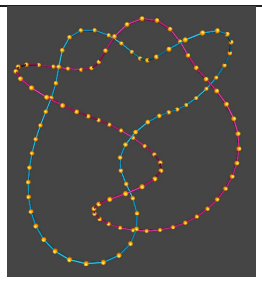
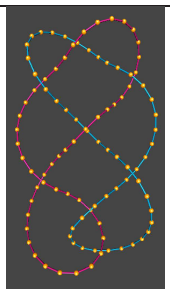
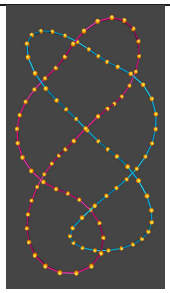
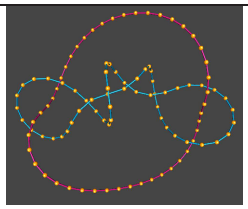
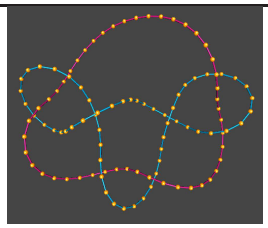
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$6_1^2 d$	7_2^+	74	100		
DLE	$6_2^2 b$	5_1^+	72	80		
DLE	$7_1^2 d$	6_1^+	70	100		
DLE	$7_4^2 a$	6_2^-	74	100		
DLE	$7_7^2 a$	8_{20}	68	100		

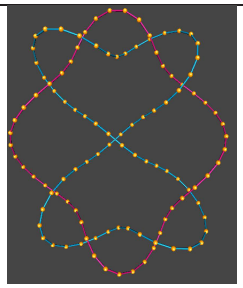
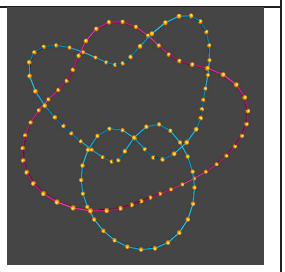
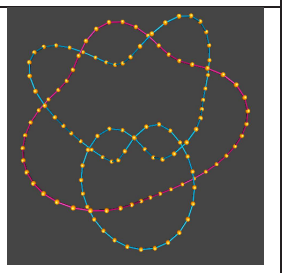
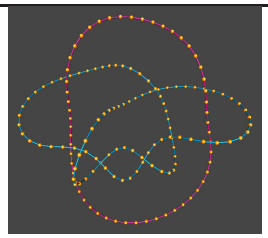
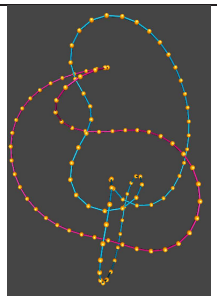
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$8_2^2 a$	9_4^+	314	–	–	
DLE	$8_2^2 c$	10_{128}^-	358	–	–	
DLE	$8_5^2 b$	$3_1^+ \# 5_2^+$	122	–	–	
DLE	$8_5^2 d$	10_{160}^-	612	–	–	
DLE	$8_6^2 d$	10_{136}^-	580	–	–	
DLE	$8_7^2 c$	$4_1 \# 4_1$	174	–	–	

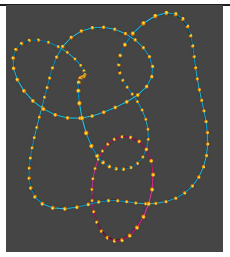
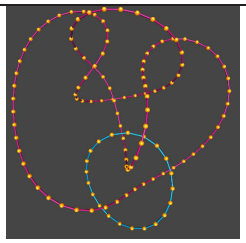
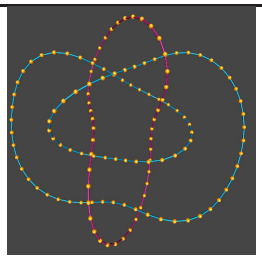
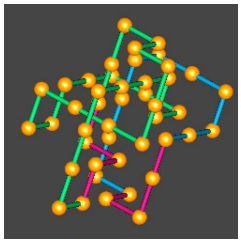
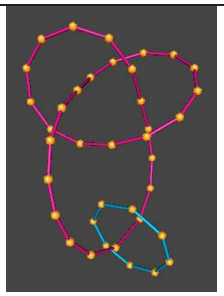
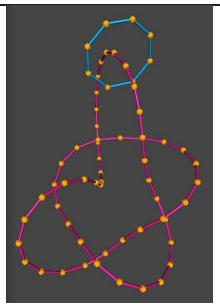
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	9_4^2c	8_4^-	292	–	–	
DLE	9_{19}^2a	9_8^-	690	–	–	
DLE	9_{19}^2b	8_5^-	324	–	–	
DLE	9_{44}^2c	7_2^-	248	–	–	
DLE	9_{51}^2b	10_{126}^+	228	–	–	

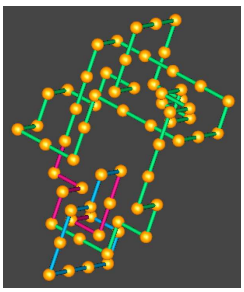
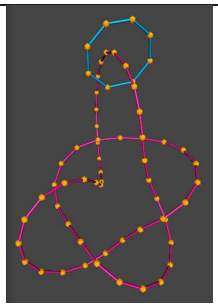
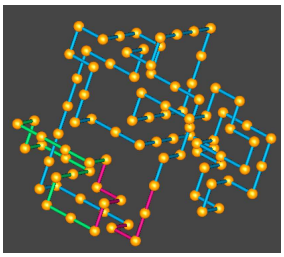
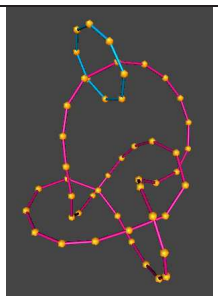
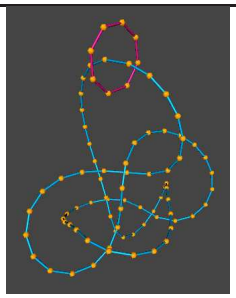
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{56}^2 a$	7_6^-	168	–	–	
DLE	$9_{57}^2 c$	8_{21}^+	120	–	–	
DLE	$9_{61}^2 b$	10_{161}^+	178	–	–	
DLE	$l(3_1^- \cup \phi)$	0_1	42	70		
DLE	$l(5_1^- \cup \phi)$	5_1^-	64	–	–	

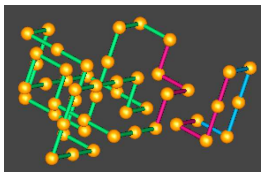
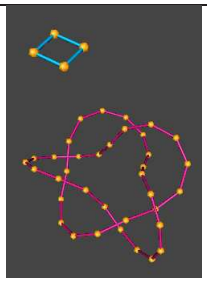
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$l(5_1^- \cup \phi)$	$3_1^+ \# 5_1^-$	92	100		
DLE	$l(5_2^- \cup \phi)$	8_{20}^-	94	100		
DLE	$l(7_3^+ \cup \phi)$	9_4^-	372	–	–	
DLE	$l(7_5^- \cup \phi)$	9_7^-	406	–	–	

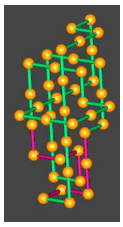
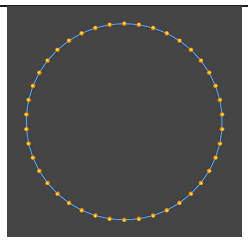
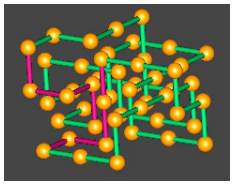
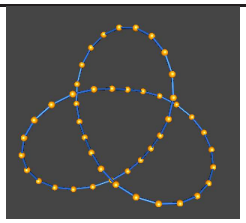
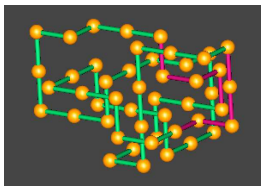
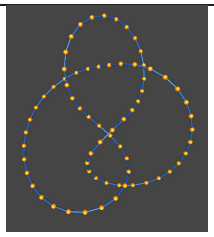
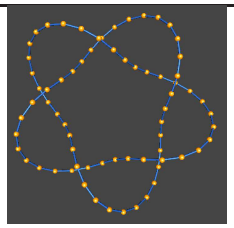
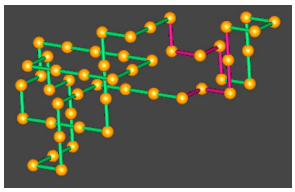
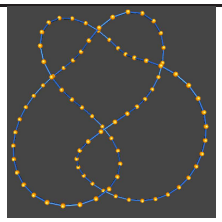
Continued on next page

Table B.6 – 5_1^- *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$u(5_1^- \cup \phi)$	5_1^-	46	100		
DLE	l_unk.	$3_1^- \# 5_1^-$	140	–	–	–
DLE	Γ_{DLE,C_1}	–	–	70	–	–
DLE	l_comp.	–	–	100	–	–
-	s	–	–	70	–	–

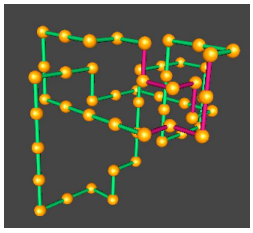
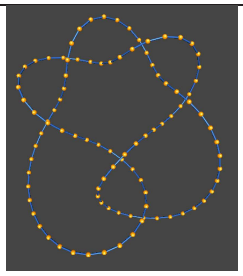
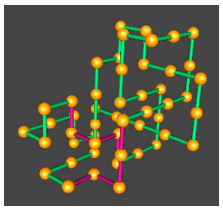
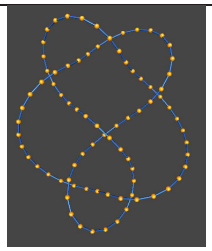
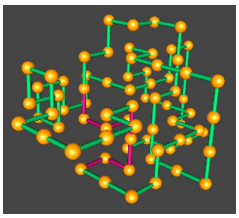
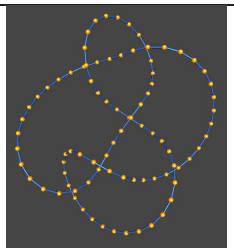
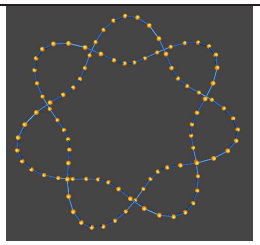
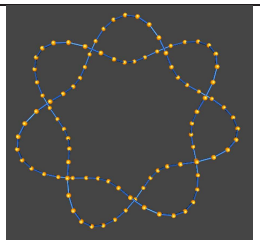
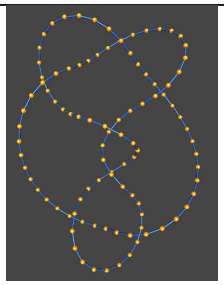
B.8 Knot Type 5_2^+ Observed Transitions

Table B.7: Table showing all observed knot and link types after a strand exchange starting with 5_2^+.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	$l(3_1^+ \cup \phi)$	42	60		
IKE	3_1^-	6_2^2a	48	80		
IKE	4_1^+	6_3^2b	46	80		
IKE	5_1^-	unk.	368	–	–	
IKE	5_2^+	$u(5_2^+ \cup \phi)$	48	80		

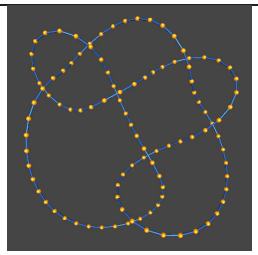
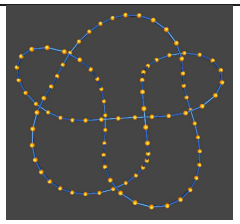
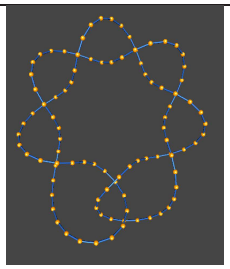
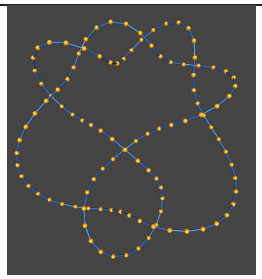
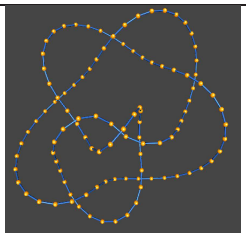
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	6_1^+	2_1^2a	50	80		
IKE	6_2^-	4_1^2d	48	80		
IKE	6_3^+	7_5^2a	72	100		
IKE	7_1^-	unk.	418	–	–	
IKE	7_5^+	9_{55}^2a	144	–	–	
IKE	7_5^-	8_4^2d	102	–	–	

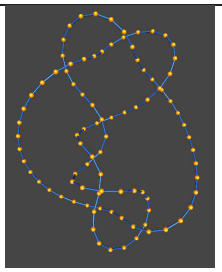
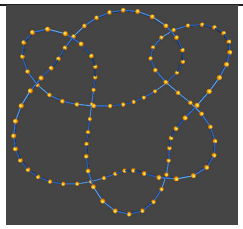
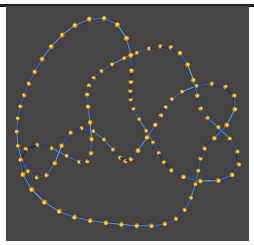
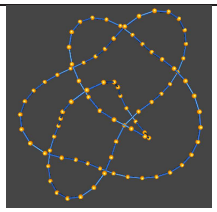
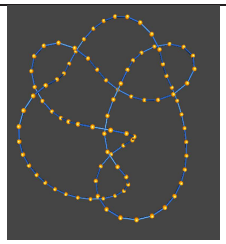
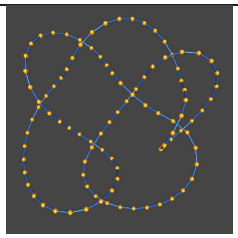
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
IKE	7_6^-	8_5^2a	122	–	–	
IKE	7_7^-	9_{58}^2c	194	–	–	
IKE	8_1^-	$n(4_1 \cup \phi)$	156	–	–	
IKE	8_2^+	$o(3_1^+ \cup \phi)$	130	–	–	
IKE	8_5^-	unk.	228	–	–	

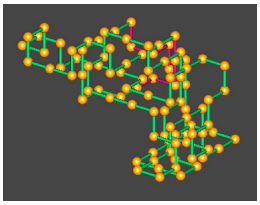
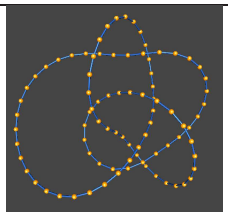
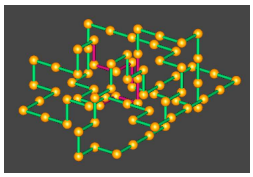
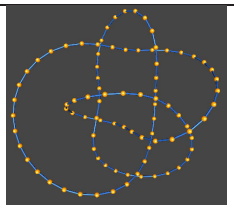
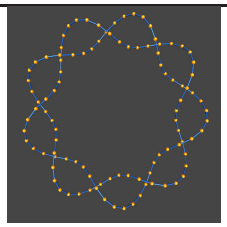
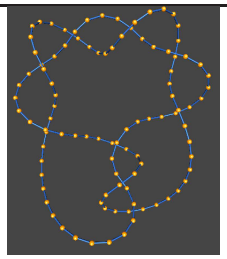
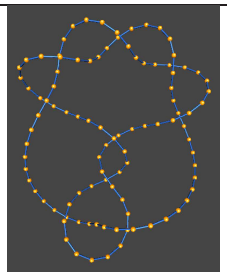
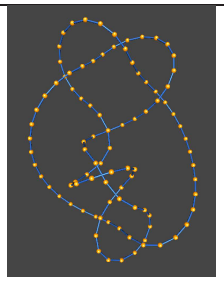
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_6^+	7_3^2a	120	–	–	
IKE	8_8^+	unk.	536	–	–	
IKE	8_9^+	7_2^2b	96	–	–	
IKE	8_{10}^-	9_{20}^2a	190	–	–	
IKE	8_{11}^-	unk.	270	–	–	
IKE	8_{14}^+	unk.	312	–	–	

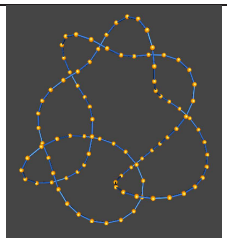
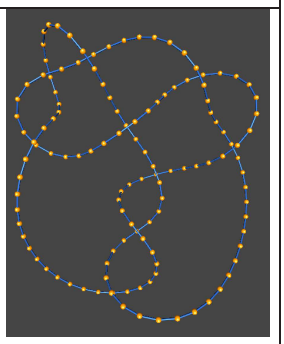
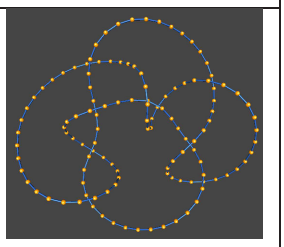
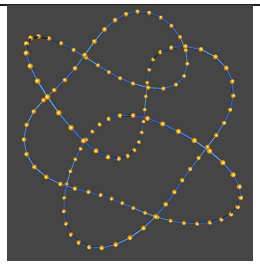
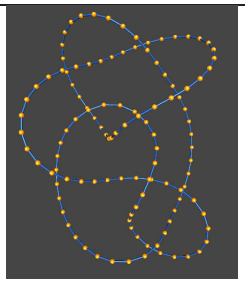
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{19}^+	$l(5_1^+ \cup \phi)$	64	80		
IKE	8_{21}^+	7_8^2a	72	100		
IKE	9_1^+	unk.	352	–	–	
IKE	9_3^-	$l(7_3^- \cup \phi)$	634	–	–	
IKE	9_6^+	$l(7_5^+ \cup \phi)$	316	–	–	
IKE	9_7^+	$l(7_2^+ \cup \phi)$	288	–	–	

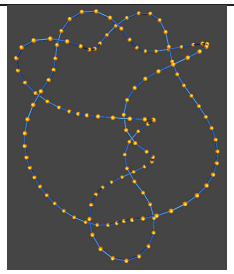
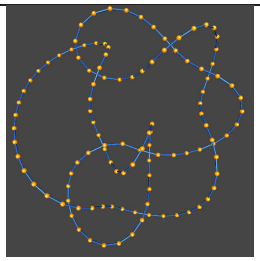
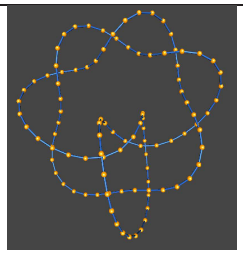
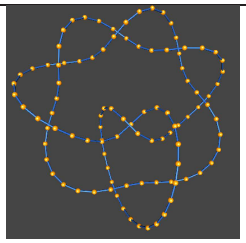
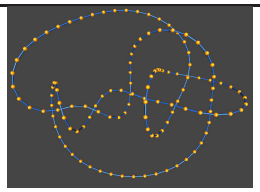
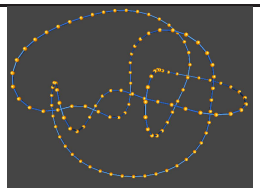
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	9_8^+	$l(7_6^+ \cup \phi)$	352	–	–	
IKE	9_{13}^-	$l(7_4^- \cup \phi)$	930	–	–	
IKE	9_{20}^+	$9_{18}^2 c$	274	–	–	
IKE	9_{42}^+	$7_1^2 a$	136	–	–	
IKE	9_{45}^+	$9_{46}^2 a$	1,266	–	–	

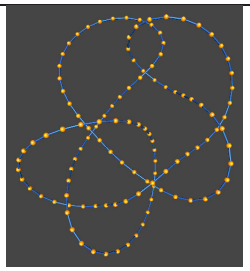
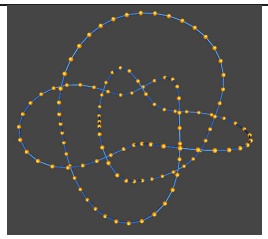
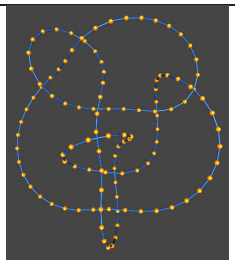
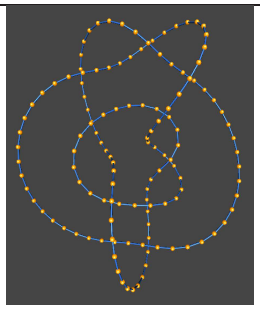
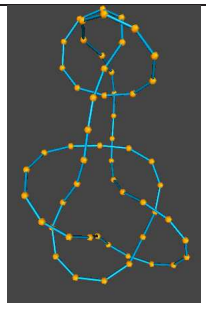
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_6^+	9_3^2a	438	–	–	
IKE	10_{54}^-	unk.	1,114	–	–	
IKE	10_{124}^+	unk.	164	–	–	
IKE	10_{125}^+	unk.	864	–	–	
IKE	10_{127}^+	9_{52}^2a	262	–	–	
IKE	10_{127}^-	8_3^2c	552	–	–	

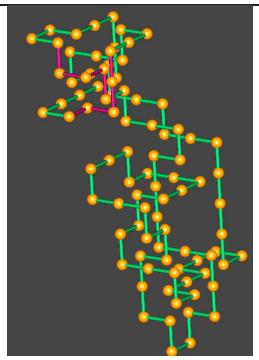
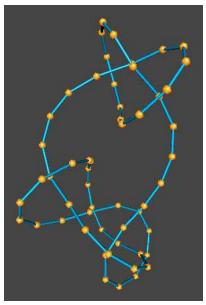
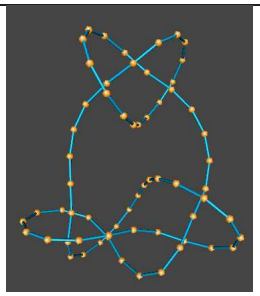
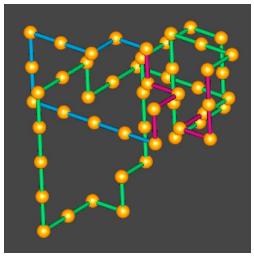
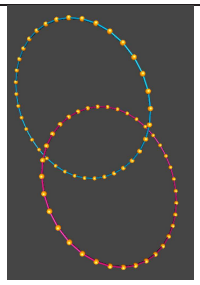
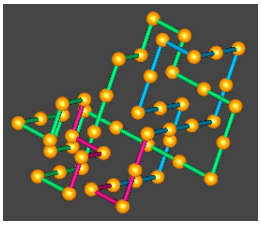
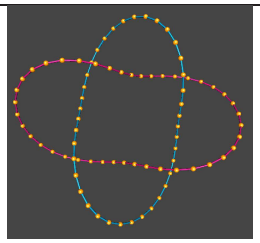
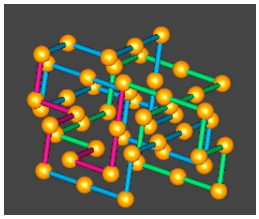
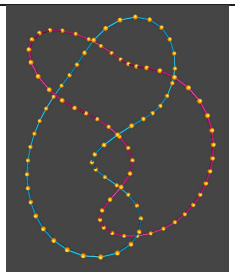
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10^-_{133}	8^2_5c	234	–	–	
IKE	10^+_{148}	unk.	396	–	–	
IKE	10^+_{151}	unk.	328	–	–	
IKE	10^+_{152}	$9^2_{59}a$	446	–	–	
IKE	$3_1^-\#5_1^-$	8^2_1d	132	–	–	

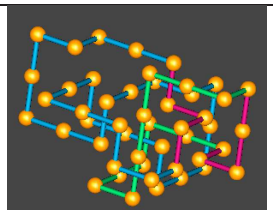
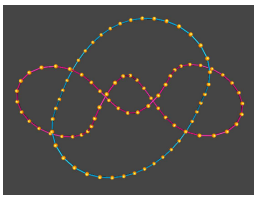
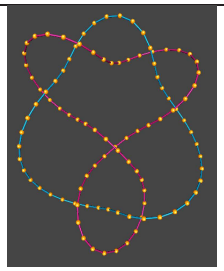
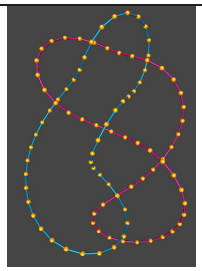
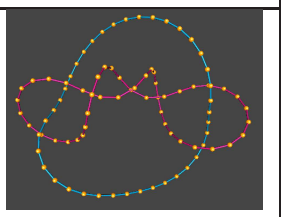
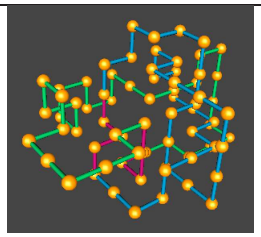
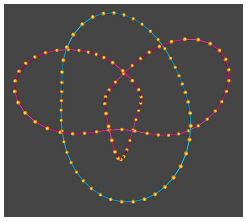
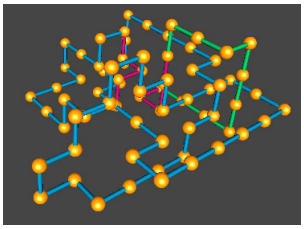
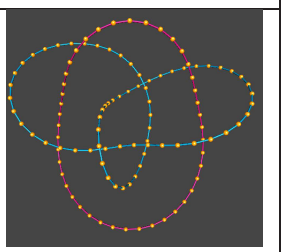
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	$3_1^+\#5_2^+$	$l(5_2^+ \cup \phi)$	88	100		
IKE	$3_1^-\#5_2^+$	unk.	142	–	–	
IKE	unk.	9_{27}^2c	246	–	–	–
IKE	Γ_{IKE,C_1}	–	–	80	–	–
IKE	comp.	–	–	100	–	–
DLE	2_1^2a	6_1^+	50	70		
DLE	4_1^2d	6_2^-	48	70		
DLE	6_2^2a	3_1^-	48	70		

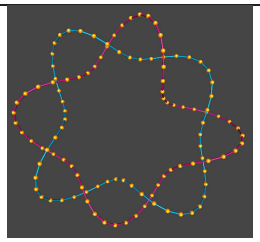
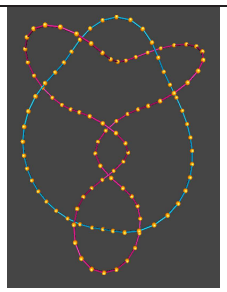
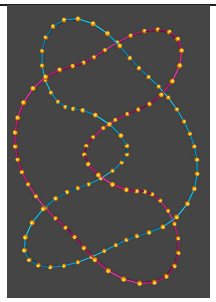
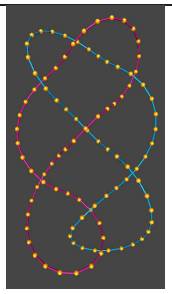
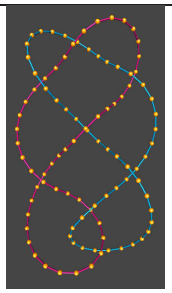
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$6_3^2 b$	4_1^+	46	80		
DLE	$7_1^2 a$	9_{42}	136	–	–	
DLE	$7_2^2 b$	8_9^+	96	–	–	
DLE	$7_3^2 a$	8_6^+	120	–	–	
DLE	$7_5^2 a$	6_3^+	72	100		
DLE	$7_8^2 a$	8_{21}^+	72	80		

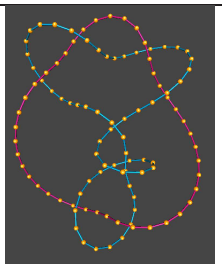
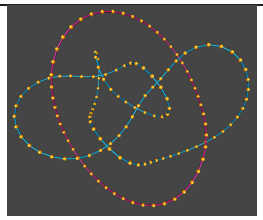
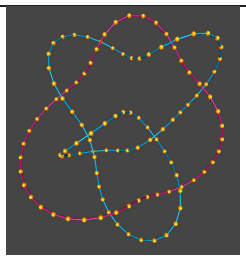
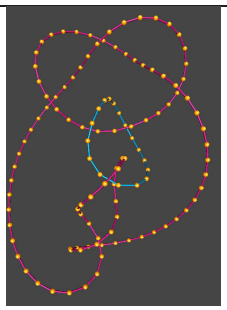
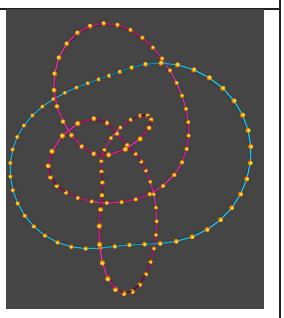
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$8_1^2 d$	$3_1^- \# 5_1^-$	132	–	–	
DLE	$8_3^2 c$	10_{127}^-	552	–	–	
DLE	$8_4^2 d$	7_5^-	102	–	–	
DLE	$8_5^2 a$	7_6^-	122	–	–	
DLE	$8_5^2 c$	10_{133}^-	234	–	–	

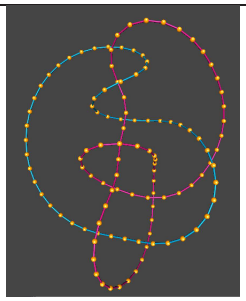
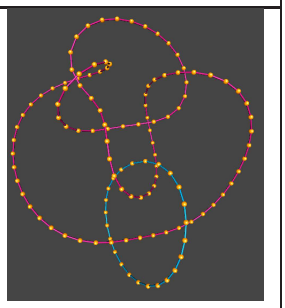
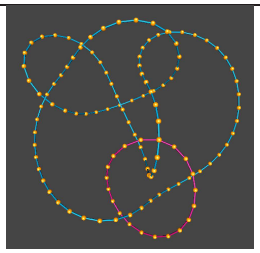
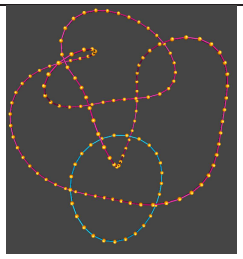
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	9_3^2a	10_6^+	438	–	–	
DLE	9_{18}^2c	9_{20}^+	274	–	–	
DLE	9_{20}^2a	8_{10}^-	190	–	–	
DLE	9_{27}^2c	unk.	246	–	–	
DLE	9_{46}^2a	9_{45}^+	1,266	–	–	

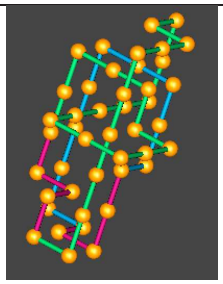
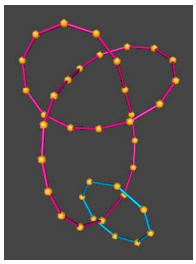
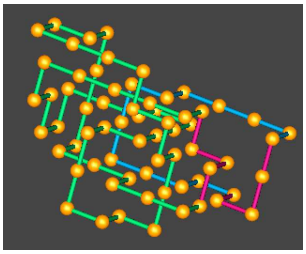
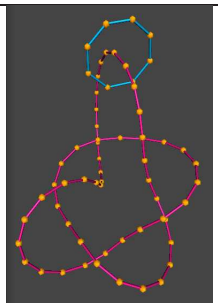
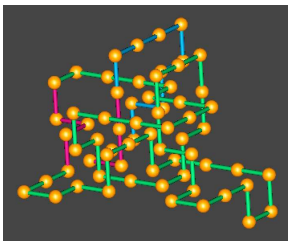
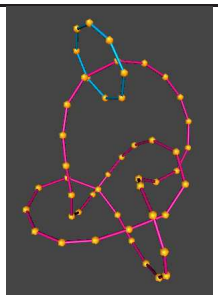
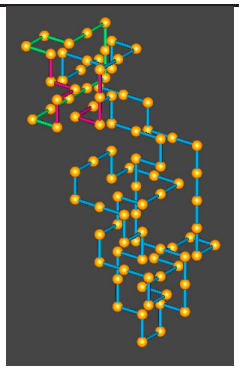
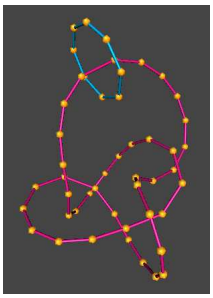
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{50}^2 a$	unk.	1,160	–	–	
DLE	$9_{52}^2 a$	10_{127}^+	262	–	–	
DLE	$9_{55}^2 a$	7_5^+	144	–	–	
DLE	$9_{58}^2 c$	7_7^-	194	–	–	
DLE	$9_{59}^2 a$	10_{152}^+	446	–	–	

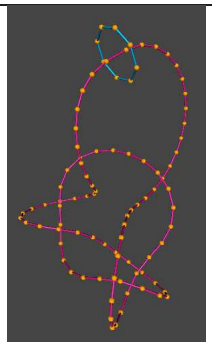
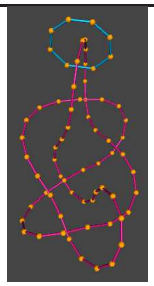
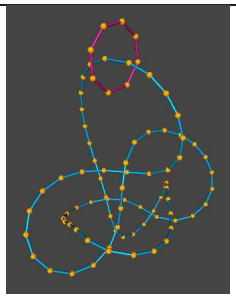
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$l(3_1^+ \cup \phi)$	0_1	42	70		
DLE	$l(5_1^+ \cup \phi)$	8_{19}^+	64	100		
DLE	$l(5_2^+ \cup \phi)$	5_2^+	64	100		
DLE	$l(5_2^+ \cup \phi)$	$3_1^+ \# 5_2^+$	88	100		

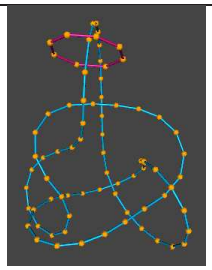
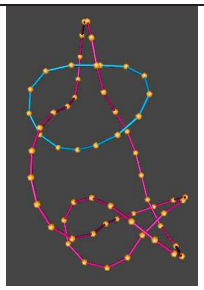
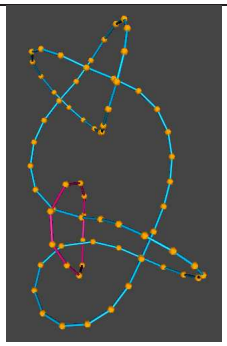
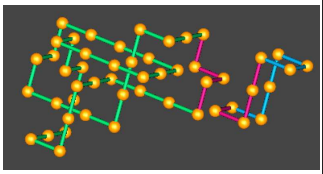
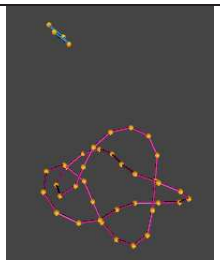
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$l(7_2^+ \cup \phi)$	9_7^+	288	–	–	
DLE	$l(7_3^- \cup \phi)$	9_3^-	634	–	–	
DLE	$l(7_4^- \cup \phi)$	9_{13}^-	930	–	–	
DLE	$l(7_5^+ \cup \phi)$	9_6^+	316	–	–	

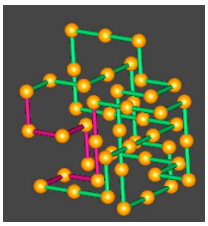
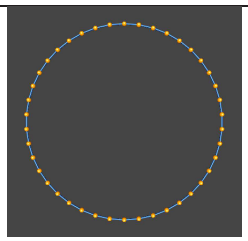
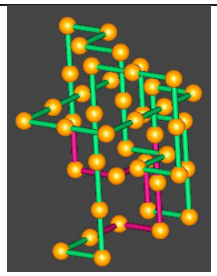
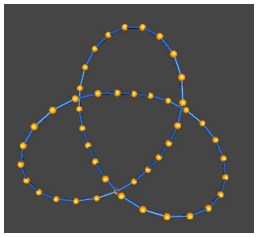
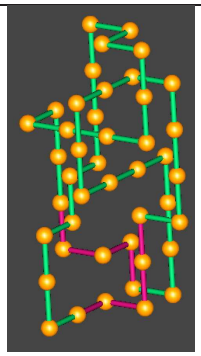
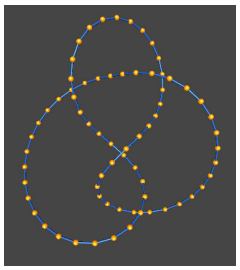
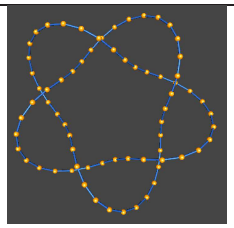
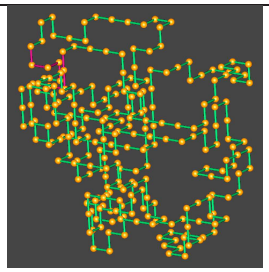
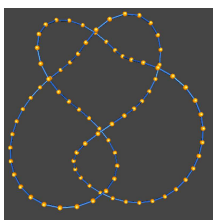
Continued on next page

Table B.7 – 5_2^+ *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(7_6^+ \cup \phi)$	9_8^+	352	–	–	
DLE	$n(4_1 \cup \phi)$	8_1^-	156	–	–	
DLE	$o(3_1^+ \cup \phi)$	8_2^+	130	–	–	
DLE	$u(5_2^+ \cup \phi)$	5_2^+	48	80		
DLE	l_unk.	$3_1^- \# 5_2^+$	142	–	–	–
DLE	Γ_{DLE,C_1}	–	–	80	–	–
DLE	l_comp.	–	–	70	–	–
-	s	–	–	80	–	–

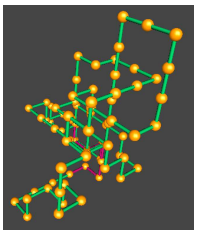
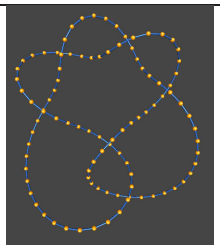
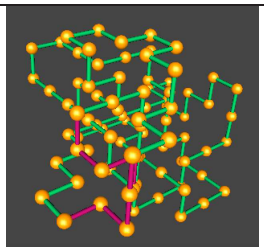
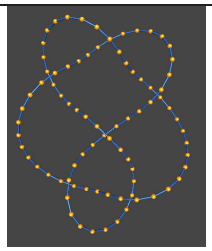
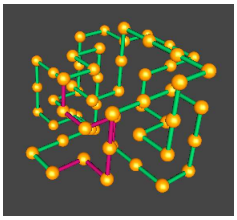
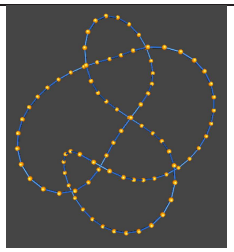
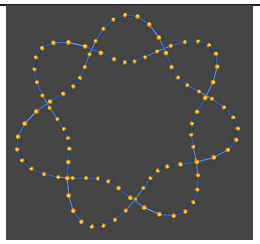
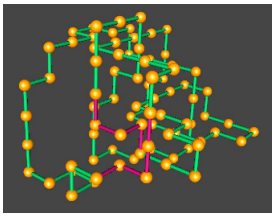
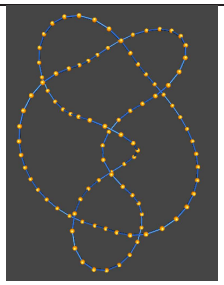
B.9 Knot Type 5_2^- Observed Transitions

Table B.8: Table showing all observed knot and link types after a strand exchange starting with 5_2^-.

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	0_1	7_8^2c	46	80		
IKE	3_1^+	4_1^2a	42	70		
IKE	4_1	2_1^2b	42	70		
IKE	5_1^+	unk.	254	–	–	
IKE	5_2^-	$u(5_2^- \cup \phi)$	46	100		

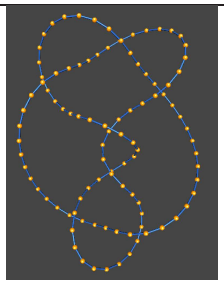
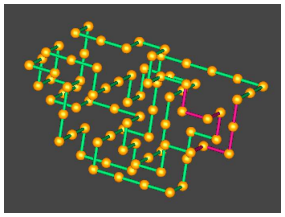
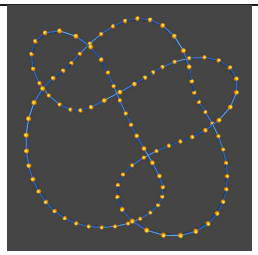
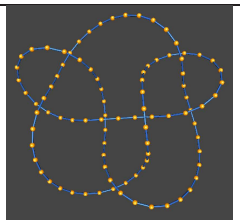
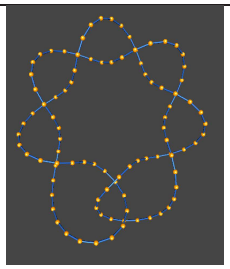
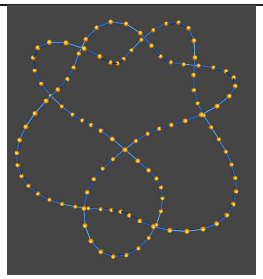
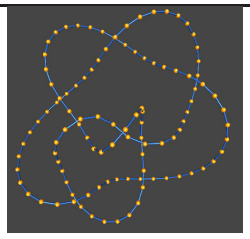
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
IKE	6_1^-	7_3^2c	80	100		
IKE	6_2^+	7_2^2c	84	100		
IKE	6_3^+	$l(3_1^- \cup \phi)$	56	80		
IKE	7_1^+	9_{51}^2d	266	–	–	
IKE	7_5^+	6_2^2b	80	100		

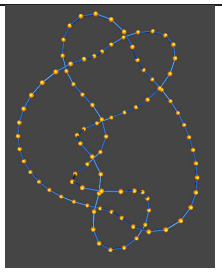
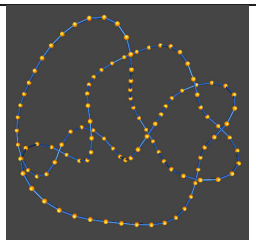
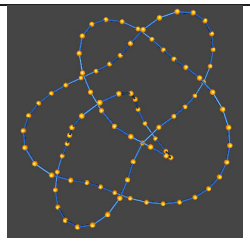
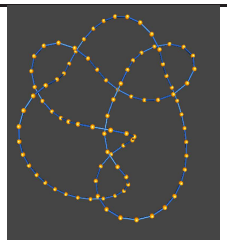
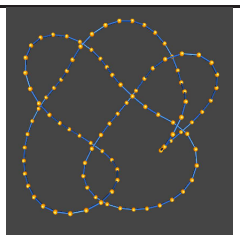
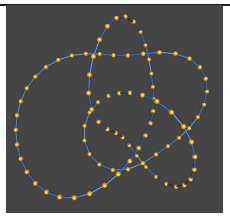
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	7_5^-	$l(5_1^- \cup \phi)$	90	–	–	
IKE	7_6^+	6_3^2c	82	100		
IKE	7_7^+	7_1^2d	90	–	–	
IKE	8_1^+	unk.	440	–	–	
IKE	8_2^-	9_{50}^2d	244	–	–	
IKE	8_5^+	8_9^2c	320	–	–	

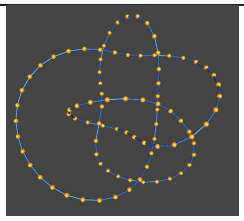
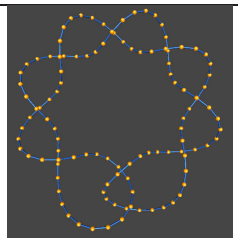
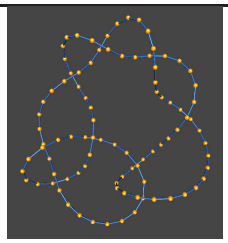
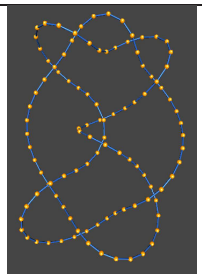
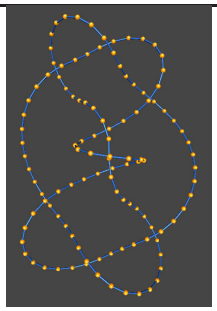
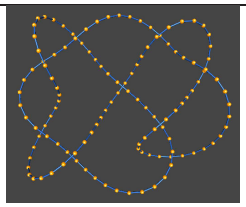
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_6^-	9_3^2d	288	–	–	
IKE	8_9^+	9_5^2b	360	–	–	
IKE	8_{10}^+	7_5^2d	134	–	–	
IKE	8_{11}^+	$n(4_1 \cup \phi)$	138	–	–	
IKE	8_{14}^-	$o(3_1^- \cup \phi)$	216	–	–	
IKE	8_{19}^-	9_{59}^2d	112	–	–	

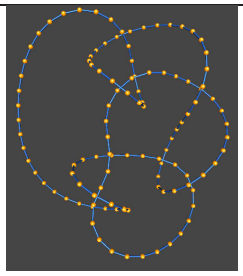
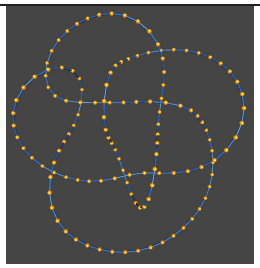
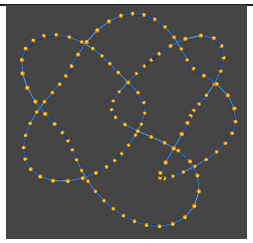
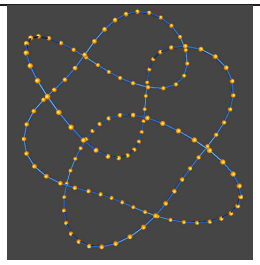
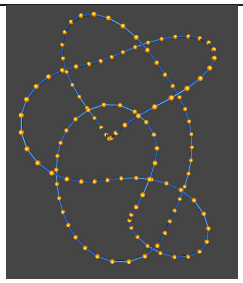
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	8_{21}^-	$9_{52}^2 d$	110	–	–	
IKE	9_2^-	$l(7_2^- \cup \phi)$	304	–	–	
IKE	9_8^-	unk.	1158	–	–	
IKE	9_9^+	$8_4^2 a$	342	–	–	
IKE	9_{10}^+	$l(7_3^+ \cup \phi)$	460	–	–	
IKE	9_{11}^-	$8_5^2 d$	214	–	–	

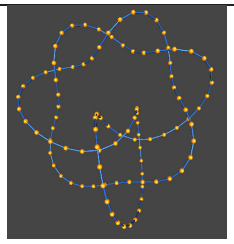
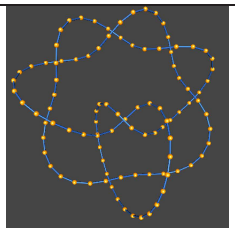
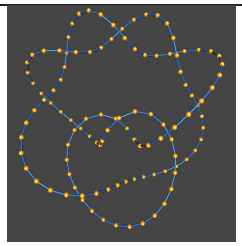
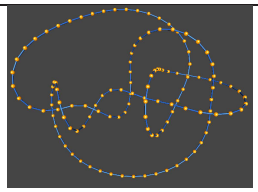
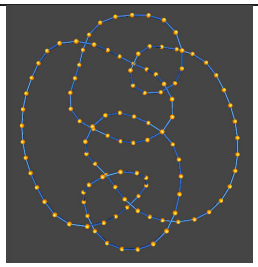
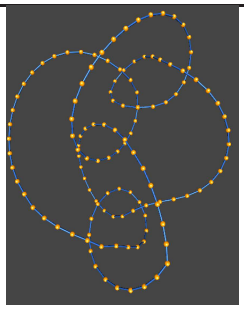
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	9_{24}^-	$l(7_6^- \cup \phi)$	908	–	–	
IKE	9_{33}^-	$9_{39}^2 a$	486	–	–	
IKE	9_{36}^-	$9_{28}^2 b$	856	–	–	
IKE	9_{42}^+	unk.	270	–	–	
IKE	9_{45}^-	unk.	360	–	–	

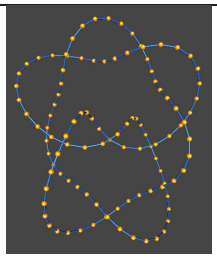
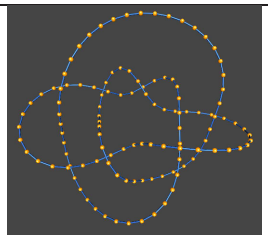
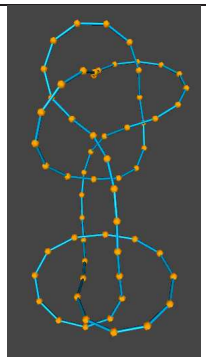
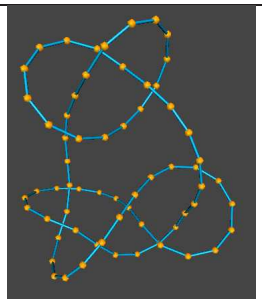
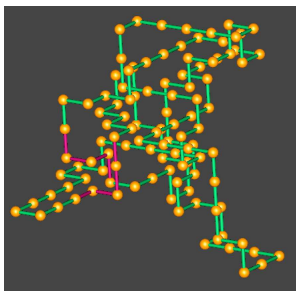
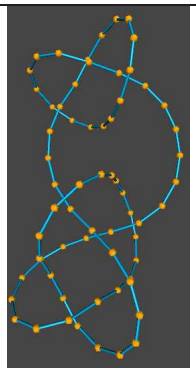
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_{124}^-	$8_1^2 a$	116	–	–	
IKE	10_{125}^+	unk.	324	–	–	
IKE	10_{126}^+	unk.	182	–	–	
IKE	10_{127}^+	unk.	816	–	–	
IKE	10_{137}^-	unk.	804	–	–	
IKE	10_{138}^-	unk.	322	–	–	

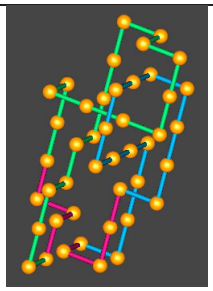
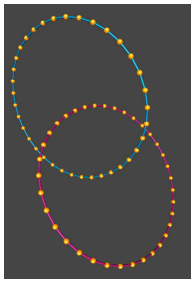
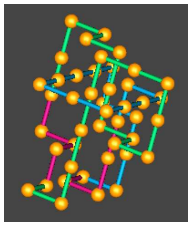
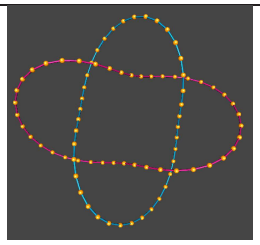
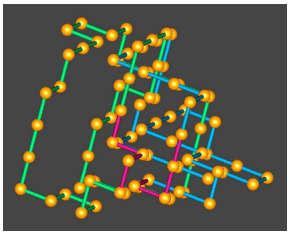
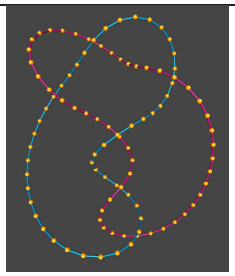
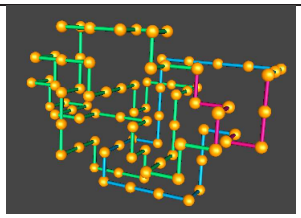
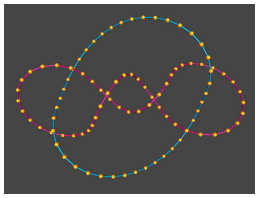
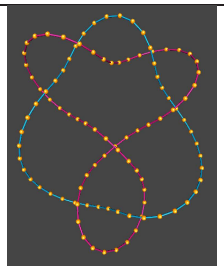
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
IKE	10_{140}^-	9_{46}^2c	166	–	–	
IKE	10_{148}^-	9_{55}^2c	240	–	–	
IKE	$3_1^+\#5_1^+$	8_3^2b	178	–	–	
IKE	$3_1^-\#5_2^-$	unk.	182	–	–	
IKE	$3_1^+\#5_2^-$	$l(5_2^- \cup \phi)$	106	100		
IKE	unk.	unk.	418	–	–	–

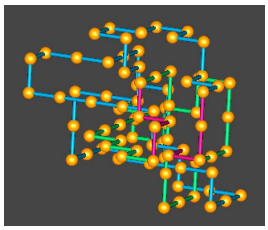
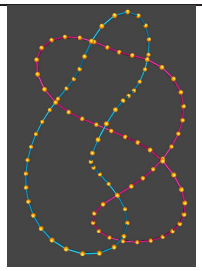
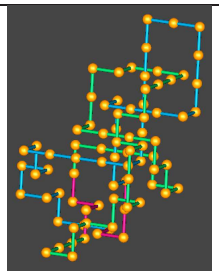
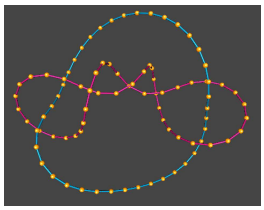
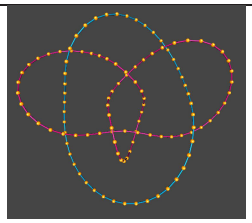
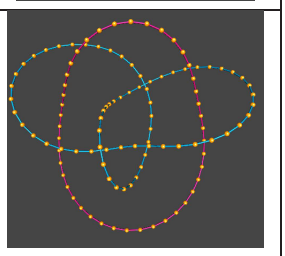
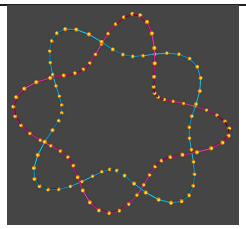
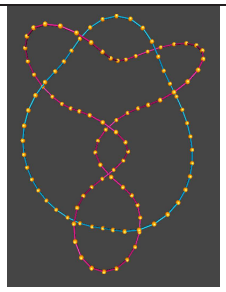
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
IKE	Γ_{IKE,C_1}	–	–	70	–	–
IKE	comp.	–	–	100	–	–
DLE	$2_1^2 b$	4_1	42	70		
DLE	$4_1^2 a$	3_1^+	42	70		
DLE	$6_2^2 b$	7_5^+	80	100		
DLE	$6_3^2 c$	7_6^+	82	100		
DLE	$7_1^2 d$	7_7^+	90	–	–	

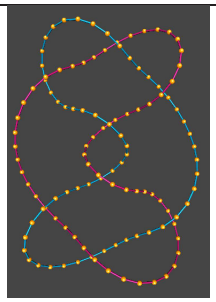
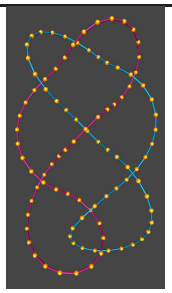
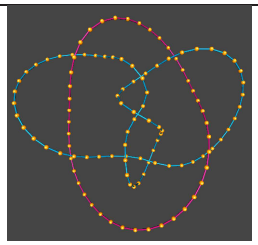
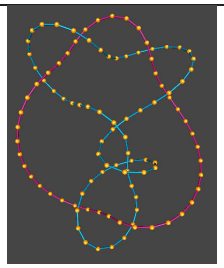
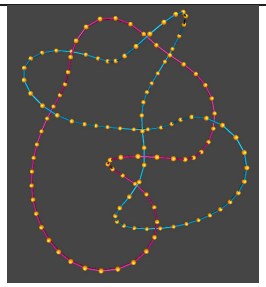
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	7_2^2c	6_2^+	84	100		
DLE	7_3^2c	6_1^-	80	100		
DLE	7_5^2d	8_{10}	134	–	–	
DLE	7_8^2c	0_1	46	80	–	
DLE	8_1^2a	10_{124}^-	116	–	–	
DLE	8_3^2b	$3_1^+\#5_1^+$	178	–	–	

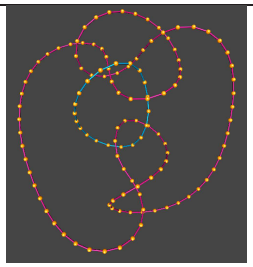
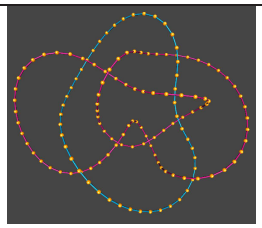
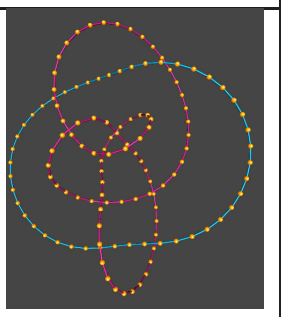
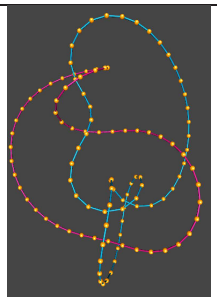
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	8_4^2a	9_9^+	342	–	–	
DLE	8_5^2d	9_{11}^-	214	–	–	
DLE	8_9^2c	8_5^+	320	–	–	
DLE	9_3^2d	8_6^-	288	–	–	
DLE	9_5^2b	8_9^+	360	–	–	

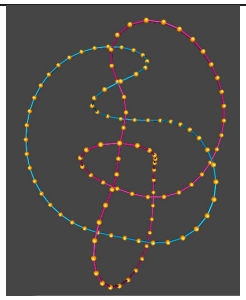
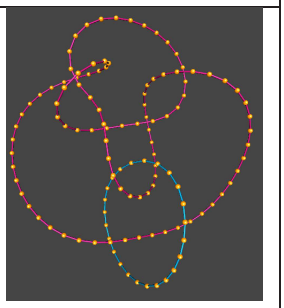
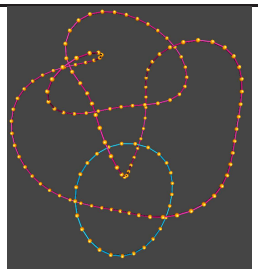
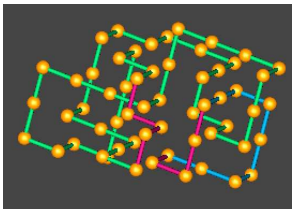
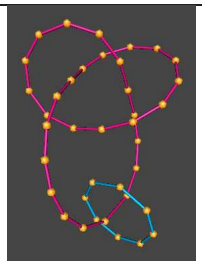
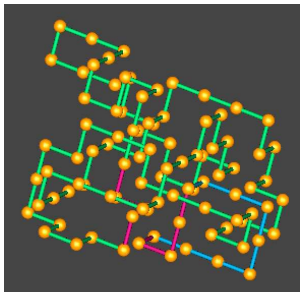
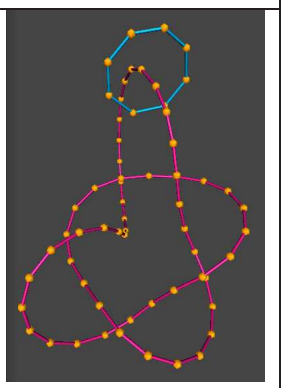
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{28}^2 b$	9_{36}^-	856	–	–	
DLE	$9_{39}^2 a$	9_{33}^-	486	–	–	
DLE	$9_{46}^2 c$	10_{140}^-	166	–	–	
DLE	$9_{50}^2 d$	8_2^-	244	–	–	
DLE	$9_{51}^2 d$	7_1^+	266	–	–	

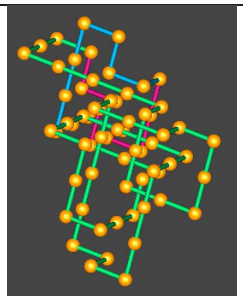
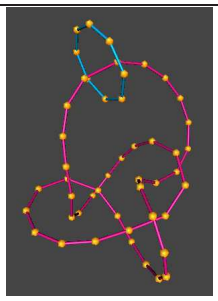
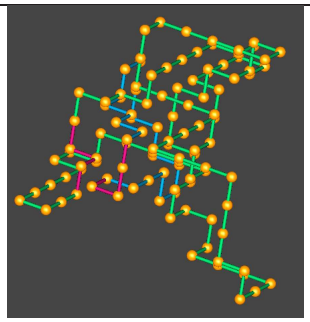
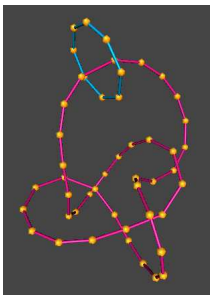
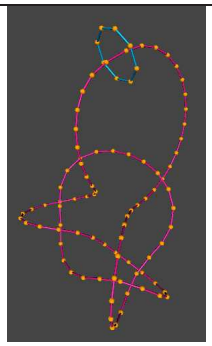
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	K/L	L/K	Size	Block Size/2	Lattice Picture	Representative
DLE	$9_{52}^2 d$	8_{21}^-	110	–	–	
DLE	$9_{55}^2 c$	10_{148}^-	240	–	–	
DLE	$9_{59}^2 d$	8_{19}^-	112	–	–	
DLE	$l(3_1^- \cup \phi)$	6_3^+	56	80		
DLE	$l(5_1^- \cup \phi)$	7_5^-	90	100		

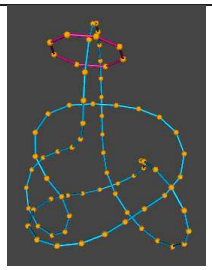
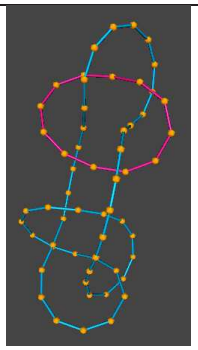
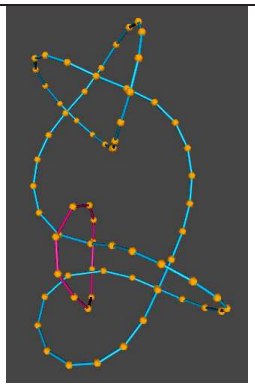
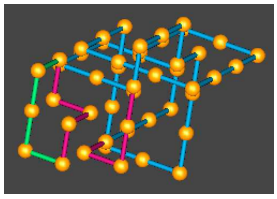
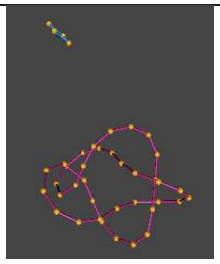
Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(5_2^- \cup \phi)$	5_2^-	62	100		
DLE	$l(5_2^- \cup \phi)$	$8c_4s$	106	100		
DLE	$l(7_2^- \cup \phi)$	9_2^-	304	–	–	
DLE	$l(7_3^+ \cup \phi)$	9_{10}^+	460	–	–	

Continued on next page

Table B.8 – 5_2^- *knot and link types after a strand exchange continued from previous page*

SE	***K/L***	***L/K***	**Size**	**Block Size/2**	**Lattice Picture**	**Representative**
DLE	$l(7_6^- \cup \phi)$	9_{24}^-	908	–	–	
DLE	$n(4_1^- \cup \phi)$	$8_{11}v$	138	–	–	
DLE	$o(3_1^- \cup \phi)$	8_{14}^-	216	–	–	
DLE	$u(5_2^- \cup \phi)$	5_2^-	46	100		
DLE	l_unk.	$3_1^- \# 5_2^-$	182	–	–	–
DLE	Γ_{DLE,C_1}	–	–	70	–	–
DLE	l_comp.	–	–	100	–	–
-	s	–	–	80	–	–

Appendix C

Appendix: Comparing LSP and LSE

C.1 Comparing LSP and LSE

This section contains Tables comparing the LSE model to the LSP model. Three tables will be presented for each knot type $K \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$: 1) a table comparing the knot type of a SP to the knot type of an IKE for knot type K; 2) a table comparing the DLE options when the SP is $K \to K$ and; 3) a table comparing the possible SP knot types for each DLE link type.

Tables C.1, C.4, C.7, C.10, C.13, C.16, C.19 and C.22 show the resulting knot types of a SP compared to the resulting knot type of an IKE for each starting knot type $K = \in \{\phi, 3_1^+, 3_1^-, 4_1, 5_1^+, 5_1^-, 5_2^+, 5_2^-\}$. In these tables the first column shows the knot type from a strand passage in the LSP model and the second column shows the knot type of an IKE in the LSE model. The third column shows the number of polygons for which this combination was found, with column four showing with the percentage out of the total $560,000,000$ sampled Θ-SAPs. Column five shows the smallest Θ-SAP size with the given combination.

Tables C.2, C.5, C.8, C.11, C.14, C.17, C.20 and C.23 show the observed link types from DLE when a SP keeps the knot type the same. Column one shows the SP knot type (the same as the starting knot), and column two shows the possible link types resulting from a DLE. The third column shows the counts of this particular combination, with the percentage of this set out of the $560,000,000$ sampled Θ-SAPs shown in column four. Column five displays the smallest knot of this combination. It can be seen that in all starting knot types K, over 99% of the links are the starting knot unlinked with an unknot, *i.e.* $u(K \cup \phi)$, with very small proportions of other links.

Tables C.3, C.6, C.9, C.12, C.15, C.18, C.21, and C.24 show the possible link types of a DLE strand exchange and the corresponding possible knot types of a SP with the LSP model. Column one shows the resulting link type of a DLE, and column two shows the resulting knot type of a SP. Column three shows the count of the number of Θ-SAP polygons with this combination, and column four shows the sample percentage of the combination is of the total $560,000,000$ sampled Θ-SAPs. Again, column five shows the smallest polygon length of such a combination.

C.2 Knot Type ϕ Tables

[Knot Type Unknot Tables]

Table C.1: Comparing SP to IKE knot type ϕ

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	$487,616,873$	87.074442	–
f	0_1	$5,847,492$	1.044195	16
f	3_1^+	$4,489$	0.000802	32
f	3_1^-	187	0.000033	66
f	5_1^+	17	0.000003	168
f	5_1^-	2	0.000000	366
f	8_{20}^-	1	0.000000	$1,196$
0_1	0_1	$65,671,007$	11.726966	14
0_1	3_1^+	6	0.000001	228
3_1^+	0_1	$836,625$	0.149397	14

Continued on next page

Table C.1 – *Comparing SP to IKE knot type ϕ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
4_1	0_1	10	0.000002	208
4_1	3_1^+	22,343	0.003990	34
5_2^+	3_1^-	831	0.000148	60
6_1^-	5_1^+	93	0.000017	138
6_2^+	5_1^+	2	0.000000	656
6_3^+	5_2^-	9	0.000002	226
7_2^+	5_1^-	10	0.000002	452
8_{20}^+	5_1^-	3	0.000001	510

Table C.2: Comparing SP to possible DLE for knot type ϕ

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
0_1	0_1^2	65,627,967	99.93445205	14
0_1	2_1^2a	214	99.93445205	36
0_1	2_1^2b	3	0.00000459	228
0_1	l_unk.	317	0.000482709	4,534

Table C.3: Comparing DLE to possible SP knot type ϕ

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	487,616,873	87.074442	–
0_1^2	f	5,655,480	1.009907	16
0_1^2	0_1	65,670,479	11.726871	14
0_1^2	3_1^+	172	0.000031	14
0_1^2	4_1	3	0.000001	208
2_1^2a	f	191,974	0.034281	22
2_1^2a	0_1	214	0.000038	36
2_1^2a	3_1^+	836,431	0.149363	24
2_1^2a	4_1	2	0.000000	404
2_1^2b	f	4,490	0.000802	32
2_1^2b	0_1	3	0.000001	228
2_1^2b	3_1^+	1	0.000000	398
2_1^2b	4_1	22,348	0.003991	34
4_1^2a	f	17	0.000003	168
4_1^2a	6_1^-	93	0.000017	138
4_1^2d	f	187	0.000033	66
4_1^2d	5_2^+	831	0.000148	60
6_1^2c	f	1	0.000000	1,064
6_1^2c	7_2^+	10	0.000002	452
7_7^2b	f	1	0.000000	366

Continued on next page

Table C.3 – *Comparing DLE to possible SP knot type ϕ*
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$7_7^2 b$	8_{20}^+	3	0.000001	510
$7_8^2 c$	f	1	0.000000	1,196
$l(3_1^+ \cup \phi)$	6_2^+	2	0.000000	656
$l(3_1^- \cup \phi)$	6_3^+	9	0.000002	226
l_unk.	f	37	0.000007	4,350
l_unk.	0_1	317	0.000057	4,534
l_unk.	3_1^+	21	0.000004	668

C.3 Knot Type 3_1^+ Tables

Table C.4: Comparing SP to IKE knot type 3_1^+

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	504,009,514	90.001699	–
f	0_1	598,103	0.106804	34
f	3_1^+	3,961,935	0.707488	32
f	4_1	441,223	0.078790	34
f	5_2^-	338	0.000060	74
f	7_1^+	1	0.000000	2,086
f	7_5^+	1	0.000000	664
f	$3_1^+ \# 3_1$	2,955	0.000528	74
f	$3_1^+ \# 3_1^-$	117	0.000021	132
f	6_2^+	350	0.000063	78
f	8_2^+	2	0.000000	714
f	$3_1^+ \# 4_1$	5	0.000001	352
f	$3_1^- \# 3_1^-$	12	0.000002	116
f	$8_{21}s$	1	0.000000	422
3_1^+	0_1	3,101	0.000554	42
3_1^+	3_1^+	44,713,664	7.984583	32
3_1^+	4_1	2,338	0.000418	60
3_1^+	5_2^-	1	0.000000	450
3_1^+	$3_1^+ \# 3_1$	16	0.000003	196
3_1^+	6_2^+	1	0.000000	1,710
5_1^+	0_1	3,277,756	0.585314	32
5_1^+	3_1^+	1,369	0.000244	44
5_1^+	4_1	89	0.000016	64
5_2^+	0_1	82	0.000015	58
5_2^+	3_1^+	982	0.000175	54
5_2^+	4_1	2,376,653	0.424402	34
7_4^+	$3_1^- \# 3_1^-$	140	0.000025	82
7_5^+	3_1^+	1	0.000000	234

Continued on next page

Table C.4 – *Comparing SP to IKE knot type 3_1^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
7_5^+	5_2^-	1,572	0.000281	70
7_6^+	6_2^+	1,661	0.000297	74
8_2^+	7_1^+	21	0.000004	208
8_{14}^+	7_5^+	2	0.000000	356
9_7^+	7_3^-	1	0.000000	1,174
9_8^+	8_2^+	14	0.000002	798
$9_4 8^-$	8_{21}^-	3	0.000001	162
10_{127}^+	7_3^+	1	0.000000	746
10_{145}^+	8_{19}^-	2	0.000000	218
10_{159}^+	7_7^+	1	0.000000	638
$3_1^+ \# 3_1^+$	0_1	49	0.000009	204
$3_1^+ \# 3_1^+$	3_1^+	590,587	0.105462	44
$3_1^+ \# 3_1^+$	4_1	33	0.000006	66
$3_1^+ \# 4_1$	0_1	3	0.000001	386
$3_1^+ \# 4_1$	3_1	12	0.000002	578
$3_1^+ \# 4_1$	4_1	1	0.000000	1,662
$3_1^+ \# 4_1$	3_1	14,770	0.002638	76
$3_1^+ \# 5_2^+$	$3_1^+ \# 3_1^-$	503	0.000090	122
unk.	$3_1^+ \# 5_1^+$	12	0.000002	296

Table C.5: Comparing SP to possible DLE for knot type 3_1^+

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
3_1	$2_1^2 a$	460	0.001029	60
3_1	$4_1^2 b$	622	0.001391	42
3_1	$u(3_1^+ \cup \phi)$	44,717,131	99.995550	32
3_1	$l(3_1^+ \cup \phi)$	2	0.000004	196
3_1	$l(3_1^+ \cup \phi)$	113	0.000253	130
3_1	l_unk.	793	0.001773	184

Table C.6: Comparing DLE to possible SP knot type 3_1^+

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	504,009,514	90.001699	–
$2_1^2 a$	f	441,170	0.078780	34
$2_1^2 a$	3_1^+	460	0.000082	60
$2_1^2 a$	5_1^+	29	0.000005	64
$2_1^2 a$	5_2^+	2,377,372	0.424531	34
$2_1^2 a$	$3_1^+ \# 3_1^+$	6	0.000001	66

Continued on next page

Table C.6 – *Comparing DLE to possible SP knot type* 3_1^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
4_1^2b	f	598,033	0.106792	34
4_1^2b	3_1^+	622	0.000111	42
4_1^2b	5_1^+	3,278,714	0.585485	32
4_1^2b	5_2^+	23	0.000004	58
4_1^2b	$3_1^+\#3_1^+$	4	0.000001	330
4_1^2b	$3_1^+\#4_1$	1	0.000000	780
5_1^2a	f	351	0.000063	78
5_1^2a	7_6^+	1,661	0.000297	74
6_1^2c	f	12	0.000002	116
6_1^2c	7_4^+	140	0.000025	82
6_2^2a	f	338	0.000060	74
6_2^2a	7_5^+	1,573	0.000281	70
6_3^2d	f	1	0.000000	422
6_3^2d	9_{48}^-	3	0.000001	162
7_1^2b	9_8^+	14	0.000002	798
7_1^2b	f	2	0.000000	714
8_2^2b	9_7^+	1	0.000000	1,174
9_{44}^2a	10_{127}^+	1	0.000000	746
9_{47}^2a	10_{133}^+	2	0.000000	670
9_{53}^2a	10_{145}^+	2	0.000000	218
9_{54}^2b	10_{159}^+	1	0.000000	638
$l(3_1^+\cup\phi)$	f	135,598	0.024214	46
$l(3_1^+\cup\phi)$	3_1^+	113	0.000020	130
$l(3_1^+\cup\phi)$	5_1^+	5	0.000001	376
$l(3_1^+\cup\phi)$	5_2^+	7	0.000001	184
$l(3_1^+\cup\phi)$	$3_1^+\#3_1^+$	590,540	0.105454	44
$l(3_1^+\cup\phi)$	f	2,957	0.000528	74
$l(3_1^+\cup\phi)$	3_1^+	2	0.000000	196
$l(3_1^+\cup\phi)$	$3_1^+\#4_1$	14,785	0.002640	76
$l(5_1^+\cup\phi)$	f	1	0.000000	2,086
$l(5_1^+\cup\phi)$	8_2^+	21	0.000004	208
$l(5_2^+\cup\phi)$	f	1	0.000000	664
$l(5_2^+\cup\phi)$	8_{14}^+	2	0.000000	356
$l(3_1^+\#3_1^+\cup\phi)$	l_unk.	2	0.000000	296
$n(3_1\cup\phi)$	f	117	0.000021	132
$n(3_1\cup\phi)$	$3_1^+\#5_2^+$	503	0.000090	122
$n(3_1^+\cup\phi)$	f	5	0.000001	352
$n(3_1^+\cup\phi)$	l_unk.	10	0.000002	300
$u(3_1^+\cup\phi)$	f	3,826,397	0.683285	32
$u(3_1^+\cup\phi)$	3_1^+	44,717,131	7.985202	32
$u(3_1^+\cup\phi)$	5_1^+	466	0.000083	44
$u(3_1^+\cup\phi)$	5_2^+	314	0.000056	56
$u(3_1^+\cup\phi)$	$3_1^+\#3_1^+$	112	0.000020	68
l_unk.	f	60	0.000011	4,640
l_unk.	3_1^+	793	0.000142	184
l_unk.	5_2^+	1	0.000000	458

Continued on next page

Table C.6 – *Comparing DLE to possible SP knot type* 3_1^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
l_unk.	$3_1^+\#3_1^+$	7	0.000001	266

C.4 Knot Type 3_1^- Tables

Table C.7: Comparing SP to IKE knot type 3_1^-

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	494,651,054	88.330545	–
f	0_1	2,684,765	0.479422	32
f	3_1^+	2,218,241	0.396114	32
f	4_1	5,432	0.000970	54
f	5_2^-	5,928	0.001059	56
f	7_3^-	5	0.000001	164
f	7_6^+	4	0.000001	616
f	8_{19}^-	14	0.000002	628
f	$3_1^+\#3_1^+$	70	0.000012	162
f	$3_1^+\#3_1^-$	1,768	0.000316	80
f	6_2^+	5	0.000001	744
f	7_2^+	2	0.000000	288
f	$3_1^+\#5_1^-$	2	0.000000	242
f	$3_1^-\#3_1^-$	12	0.000002	342
0_1	0_1	34,975,622	6.245647	30
0_1	3_1^+	74,803	0.013358	36
0_1	4_1	82	0.000015	74
0_1	5_2^-	93	0.000017	98
0_1	$3_1^+\#3_1^-$	21	0.000004	150
0_1	$3_1^-\#3_1^-$	1	0.000000	190
3_1^+	0_1	103,513	0.018484	32
3_1^+	3_1^+	24,860,942	4.439454	32
3_1^+	4_1	76	0.000014	86
3_1^+	5_2^-	72	0.000013	74
3_1^+	$3_1^+\#3_1^-$	22	0.000004	172
6_2^+	0_1	118	0.000021	40
6_2^+	3_1^+	67	0.000012	62
6_2^+	5_2^-	29,979	0.005353	52
6_3^+	0_1^+	86	0.000015	50
6_3^+	3_1^+	33	0.000006	82
6_3^+	4_1	27,075	0.004835	54
7_7^+	0_1	3	0.000001	152
7_7^+	3_1^+	1	0.000000	290
7_7^+	$3_1^-\#3_1^-$	89	0.000016	130

Continued on next page

Table C.7 – *Comparing SP to IKE knot type 3_1^- continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
8_6^+	7_3^-	23	0.000004	178
8_7^+	7_3^+	2	0.000000	454
8_8^-	6_2^+	63	0.000011	190
8_8^+	7_2^+	7	0.000001	208
8_{16}^+	7_6^+	14	0.000002	350
9_{46}^+	8_{19}^-	43	0.000008	164
10_{125}^+	7_1^+	1	0.000000	318
10_{135}^-	7_5^+	1	0.000000	3,234
10_{146}^-	8_{21}^-	1	0.000000	1,796
$3_1^+\#3_1^-$	0_1	1,451	0.000259	42
$3_1^+\#3_1^-$	3_1^+	348,709	0.062269	46
$3_1^+\#3_1^-$	4_1	1	0.000000	306
$3_1^+\#3_1^-$	5_2^-	1	0.000000	436
$3_1^+\#4_1$	0_1	25	0.000004	74
$3_1^+\#4_1$	3_1^+	12	0.000002	62
$3_1^+\#4_1$	$3_1^+\#3_1^-$	9,321	0.001664	72
$3_1^+\#5_2^-$	0_1	4	0.000001	70
$3_1^+\#5_2^-$	$3_1^+\#3_1^+$	310	0.000055	140
unk.	$3_1^+\#5_1^-$	11	0.000002	172

Table C.8: Comparing SP to possible DLE for knot type 3_1^-

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
3_1^+	2_1^2a	92,983	0.372459	32
3_1^+	$u(3_1^+ \cup \phi)$	24,870,659	99.623603	32
3_1^+	$l(3_1^+ \cup \phi)$	614	0.002459	70
3_1^+	$l(3_1^+ \cup \phi)$	9	0.000036	172
3_1^+	4_1^2b	64	0.000256	74
3_1^+	5_1^2a	69	0.000276	86
3_1^+	l_unk.	227	0.000909	296

Table C.9: Comparing DLE to possible SP knot type 3_1^-

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	494,651,054	88.330545	–
2_1^2a	f	2,683,999	0.479286	32
2_1^2a	0_1	34,978,517	6.246164	30
2_1^2a	3_1^+	92,983	0.016604	32
2_1^2a	6_2^+	109	0.000019	40

Continued on next page

Table C.9 – *Comparing DLE to possible SP knot type* 3_1^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
2_1^2a	6_3^+	81	0.000014	50
2_1^2a	7_7^+	3	0.000001	152
2_1^2a	$3_1^+\#3_1^-$	1,262	0.000225	42
2_1^2a	$3_1^+\#4_1$	23	0.000004	74
2_1^2a	$3_1^+\#5_2^-$	4	0.000001	70
4_1^2b	f	5,929	0.001059	56
4_1^2b	0_1	89	0.000016	98
4_1^2b	3_1^+	64	0.000011	74
4_1^2b	6_2^+	29,992	0.005356	52
4_1^2b	$3_1^+\#3_1^-$	1	0.000000	436
5_1^2a	f	5,432	0.000970	54
5_1^2a	0_1	80	0.000014	74
5_1^2a	3_1^+	69	0.000012	86
5_1^2a	6_3^+	27,081	0.004836	54
5_1^2a	$3_1^+\#3_1^-$	1	0.000000	306
6_1^2c	f	14	0.000002	628
6_1^2c	9_{46}^+	43	0.000008	164
6_2^2a	f	5	0.000001	164
6_2^2a	8_6^+	23	0.000004	178
6_3^2d	f	12	0.000002	342
6_3^2d	0_1	1	0.000000	190
6_3^2d	7_7^+	89	0.000016	130
7_1^2b	f	5	0.000001	744
7_1^2b	8_8^-	63	0.000011	190
7_4^2c	f	4	0.000001	616
7_4^2c	8_{16}^+	14	0.000002	350
9_{43}^2d	10_{125}^+	1	0.000000	318
9_{48}^2c	10_{135}^-	1	0.000000	3,234
9_{54}^2d	10_{146}^-	1	0.000000	1,796
$l(3_1^+\cup\phi)$	f	1,770	0.000316	80
$l(3_1^+\cup\phi)$	0_1	20	0.000004	150
$l(3_1^+\cup\phi)$	3_1^+	9	0.000002	172
$l(3_1^+\cup\phi)$	$3_1^+\#4_1$	9,324	0.001665	72
$l(3_1^+\cup\phi)$	f	79,213	0.014145	44
$l(3_1^+\cup\phi)$	0_1	772	0.000138	78
$l(3_1^+\cup\phi)$	3_1^+	614	0.000110	70
$l(3_1^+\cup\phi)$	$3_1^+\#3_1^-$	348,167	0.062173	46
$l(5_1^+\cup\phi)$	8_7^+	2	0.000000	454
$l(5_2^+\cup\phi)$	f	2	0.000000	288
$l(5_2^+\cup\phi)$	8_8^+	7	0.000001	208
$n(3_1\cup\phi)$	f	1	0.000000	242
$n(3_1\cup\phi)$	l_unk.	10	0.000002	172
$n(3_1^+\cup\phi)$	f	70	0.000012	162
$n(3_1^+\cup\phi)$	$3_1^+\#5_2^-$	310	0.000055	140
$u(3_1^+\cup\phi)$	f	2,139,776	0.382103	32
$u(3_1^+\cup\phi)$	0_1	71,140	0.012704	36
$u(3_1^+\cup\phi)$	3_1^+	24,870,659	4.441189	32

Continued on next page

Table C.9 – *Comparing DLE to possible SP knot type* 3_1^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$u(3_1^+ \cup \phi)$	6_2^+	63	0.000011	62
$u(3_1^+ \cup \phi)$	6_3^+	32	0.000006	82
$u(3_1^+ \cup \phi)$	7_7^+	1	0.000000	290
$u(3_1^+ \cup \phi)$	$3_1^+ \# 3_1^-$	725	0.000129	50
$u(3_1^+ \cup \phi)$	$3_1^+ \# 4_1$	11	0.000002	62
l_unk.	f	15	0.000003	6,690
l_unk.	0_1	3	0.000001	182
l_unk.	3_1^+	227	0.000041	296
l_unk.	$3_1^+ \# 3_1^-$	6	0.000001	7,230

C.5 Knot Type 4_1 Tables

Table C.10: Comparing SP to IKE knot type 4_1

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	493,790,078	88.176800	–
f	3_1^+	174,844	0.031222	44
f	3_1^-	28,958,912	5.171234	36
f	4_1	2,778,190	0.496105	42
f	5_2^+	250,801	0.044786	44
f	5_2^-	5,374	0.000960	68
f	6_3^+	4,245	0.000758	72
f	7_1^+	27	0.000005	196
f	7_1^-	1	0.000000	306
f	7_4^+	1	0.000000	668
f	7_5^+	176	0.000031	120
f	7_5^-	19	0.000003	148
f	$3_1^+ \# 4_1$	2,045	0.000365	80
f	$3_1^- \# 4_1$	85	0.000015	216
f	8_1^-	2	0.000000	396
f	8_7^+	9	0.000002	268
f	9_6^+	1	0.000000	1,420
f	10_{143}^-	2	0.000000	420
f	l_unk.	2	0.000000	238
4_1	3_1^-	32,856	0.005867	50
4_1	3_1^+	1,441	0.000257	78
4_1	4_1	31,048,787	5.544426	40
4_1	5_2^+	2,026	0.000362	72
4_1	5_2^-	34	0.000006	122
4_1	6_3^+	28	0.000005	346
4_1	7_1^+	1	0.000000	400

Continued on next page

Table C.10 – *Comparing SP to IKE knot type* 4_1 *continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
4_1	7_5^+	1	0.000000	3,506
4_1	$3_1^+\#4_1$	3	0.000001	700
unk.	l_unk.	22	0.000004	276
unk.	$10_{139}s$	5	0.000001	312
6_1^+	3_1^+	28	0.000005	74
6_1^+	3_1^-	651	0.000116	58
6_1^+	4_1	977	0.000174	64
6_1^+	5_2^+	1,489,128	0.265916	42
6_1^+	5_2^-	2	0.000000	386
6_1^+	7_5^+	1	0.000000	682
6_2^+	3_1^+	975,505	0.174197	44
6_2^+	3_1^-	526	0.000094	80
6_2^+	4_1	715	0.000128	52
6_2^+	5_2^+	31	0.000006	74
6_2^+	6_3^+	4	0.000001	2,454
6_2^+	$3_1^+\#4_1$	1	0.000000	1,026
7_6^+	3_1^+	3	0.000001	712
7_6^+	3_1^-	18	0.000003	114
7_6^+	4_1	23	0.000004	142
7_6^+	5_2^+	2	0.000000	336
7_6^+	5_2^-	27,279	0.004871	74
8_{12}^+	3_1^-	1	0.000000	452
8_{12}^+	7_5^+	833	0.000149	134
7_7^+	3_1^-	6	0.000001	384
7_7^+	4_1	14	0.000002	192
7_7^+	6_3^+	20,033	0.003577	66
8_4^+	7_1^+	162	0.000029	100
8_{13}^+	7_5^-	64	0.000011	130
8_{14}^+	6_3^+	519	0.000093	132
8_{16}^+	7_4^-	6	0.000001	196
9_8^+	8_1^-	4	0.000001	196
9_{15}^-	7_5^-	37	0.000007	186
9_12^+	$3_1^-\#4_1$	417	0.000074	176
9_{17}^+	8_2^+	6	0.000001	152
9_{19}^+	8_6^+	3	0.000001	292
9_{19}^-	8_7^+	76	0.000014	218
9_{27}^+	8_{11}^-	2	0.000000	336
9_{33}^+	8_{15}^-	1	0.000000	508
10_{35}^+	9_6^+	14	0.000002	1,084
10_{36}^+	8_7^-	1	0.000000	394
10_{140}^+	7_1^-	8	0.000001	180
10_{144}^+	3_1^+	1	0.000000	756
10_{144}^+	7_5^+	6	0.000001	118
10_{155}^-	7_4^+	2	0.000000	296
10_{158}^-	9_{49}^-	3	0.000001	566

Continued on next page

Table C.10 – *Comparing SP to IKE knot type* 4_1 *continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
10^-_{163}	9^-_{45}	2	0.000000	372
$3_1^+\#4_1$	3_1^+	19	0.000003	96
$3_1^+\#4_1$	3_1^-	610	0.000109	68
$3_1^+\#4_1$	4_1	421,900	0.075339	66
$3_1^+\#4_1$	5_2^+	36	0.000006	140
$3_1^+\#4_1$	5_2^-	2	0.000000	822
$3_1^+\#4_1$	$3_1^+\#4_1$	1	0.000000	3,770
$4_1\#4_1$	3_1^-	15	0.000003	200
$4_1\#4_1$	4_1	21	0.000004	800
$4_1\#4_1$	5_2^+	1	0.000000	458
$4_1\#4_1$	$3_1^+\#4_1$	10,262	0.001833	92

Table C.11: Comparing SP to possible DLE for knot type 4_1

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
4_1	2_1^2a	31,929	0.102715	50
4_1	2_1^2b	2,094	0.006736	72
4_1	5_1^2a	1,388	0.004465	78
4_1	5_1^2c	26	0.000084	308
4_1	6_3^2a	2	0.000006	248
4_1	6_3^2d	25	0.000080	122
4_1	$l(4_1 \cup \phi)$	660	0.002123	74
4_1	$l(4_1 \cup \phi)$	19	0.000061	660
4_1	$n(3_1^+ \cup \phi)$	1	0.000003	400
4_1	$u(4_1\ cup\phi)$	31,047,917	99.880136	40
4_1	l_unk.	1,116	0.003590	188

Table C.12: Comparing DLE to possible SP knot type 4_1

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	493,790,078	88.176800	–
2_1^2a	f	2,098,984	0.374819	38
2_1^2a	0_1	26,861,300	4.796661	36
2_1^2a	4_1	31,929	0.005702	50
2_1^2a	6_1^+	653	0.000117	58
2_1^2a	6_2^+	467	0.000083	80
2_1^2a	7_6^+	14	0.000002	114
2_1^2a	7_7^+	9	0.000002	216
2_1^2a	$3_1^+\#4_1$	541	0.000097	68

Continued on next page

Table C.12 – *Comparing DLE to possible SP knot type* 4_1
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
2_1^2a	$4_1\#4_1$	9	0.000002	200
2_1^2b	f	250,131	0.044666	44
2_1^2b	0_1	638	0.000114	62
2_1^2b	4_1	2,094	0.000374	72
2_1^2b	6_1^+	1,489,054	0.265903	42
2_1^2b	6_2^+	32	0.000006	74
2_1^2b	$3_1^+\#4_1$	34	0.000006	140
5_1^2a	f	174,376	0.031139	44
5_1^2a	0_1	437	0.000078	62
5_1^2a	4_1	1,388	0.000248	78
5_1^2a	6_1^+	32	0.000006	74
5_1^2a	6_2^+	975,584	0.174211	44
5_1^2a	8_{14}^+	1	0.000000	862
5_1^2a	$3_1^+\#4_1$	20	0.000004	96
5_1^2c	f	4,133	0.000738	72
5_1^2c	0_1	10	0.000002	126
5_1^2c	4_1	26	0.000005	308
5_1^2c	6_2^+	2	0.000000	2,454
5_1^2c	7_7^+	20,033	0.003577	66
6_3^2a	f	173	0.000031	120
6_3^2a	4_1	2	0.000000	248
6_3^2a	8_{12}	830	0.000148	134
6_3^2d	f	5,359	0.000957	68
6_3^2d	0_1	9	0.000002	108
6_3^2d	4_1	25	0.000004	122
6_3^2d	6_1^+	1	0.000000	622
6_3^2d	7_6^+	27,295	0.004874	74
6_3^2d	$3_1^+\#4_1$	1	0.000000	822
7_2^2b	f	104	0.000019	130
7_2^2b	6_2^+	1	0.000000	1,028
7_2^2b	8_{14}^+	518	0.000092	132
7_2^2c	f	9	0.000002	268
7_2^2c	9_{19}^-	76	0.000014	218
7_5^2b	10_{158}^-	3	0.000001	566
7_5^2c	8_{16}^+	6	0.000001	196
8_3^2b	f	1	0.000000	1,420
8_3^2b	10_{35}^+	14	0.000002	1,084
8_3^2c	f	8	0.000001	326
8_3^2c	9_{15}^-	37	0.000007	186
8_{11}^2b	9_{33}^+	1	0.000000	508
9_2^2a	10_{36}^+	1	0.000000	394
9_{49}^2a	f	1	0.000000	306
9_{49}^2a	10_{140}^+	8	0.000001	180
9_{49}^2d	l_unk.	5	0.000001	312
9_{50}^2a	f	1	0.000000	668
9_{50}^2a	10_{155}^-	2	0.000000	296
9_{52}^2a	f	2	0.000000	1,326

Continued on next page

Table C.12 – *Comparing DLE to possible SP knot type* 4_1
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$9^2_{52}a$	10^+_{144}	7	0.000001	118
$9^2_{52}d$	f	2	0.000000	420
$9^2_{58}a$	10^-_{163}	2	0.000000	372
$l(4_1 \cup \phi)$	f	96,496	0.017231	54
$l(4_1 \cup \phi)$	0_1	238	0.000043	64
$l(4_1 \cup \phi)$	4_1	660	0.000118	74
$l(4_1 \cup \phi)$	6^+_1	17	0.000003	224
$l(4_1 \cup \phi)$	6^+_2	8	0.000001	380
$l(4_1 \cup \phi)$	$3^+_1\#4_1$	421156	0.075206	66
$l(4_1 \cup \phi)$	f	2,044	0.000365	80
$l(4_1 \cup \phi)$	0_1	8	0.000001	150
$l(4_1 \cup \phi)$	4_1	19	0.000003	660
$l(4_1 \cup \phi)$	$3^+_1\#4_1$	1	0.000000	3,770
$l(4_1 \cup \phi)$	$4_1\#4_1$	10,271	0.001834	92
$l(4_1 \cup \phi)$	9^+_{19}	3	0.000001	292
$l(6^-_1 \cup \phi)$	f	2	0.000000	396
$l(6^-_1 \cup \phi)$	9^+_8	4	0.000001	196
$l(6^+_2 \cup \phi)$	9^+_{17}	6	0.000001	152
$l(6^-_2 \cup \phi)$	9^+_{27}	2	0.000000	336
$l(3^+_1\#4_1 \cup \phi)$	l_unk.	2	0.000000	720
$n(3^+_1 \cup \phi)$	f	26	0.000005	196
$n(3^+_1 \cup \phi)$	4_1	1	0.000000	400
$n(3^+_1 \cup \phi)$	8^+_4	162	0.000029	100
$n(3^-_1 \cup \phi)$	f	11	0.000002	148
$n(3^-_1 \cup \phi)$	8^+_{13}	64	0.000011	130
$n(4_1 \cup \phi)$	f	2	0.000000	238
$n(4_1 \cup \phi)$	l_unk.	19	0.000003	276
$n(4_1 \cup \phi)$	f	86	0.000015	216
$n(4_1 \cup \phi)$	9^+_{12}	417	0.000074	176
$u(4_1 \cup \phi)$	f	2,665,173	0.475924	42
$u(4_1 \cup \phi)$	0_1	14,750	0.002634	48
$u(4_1 \cup \phi)$	4_1	31,047,917	5.544271	40
$u(4_1 \cup \phi)$	6^+_1	1,030	0.000184	74
$u(4_1 \cup \phi)$	6^+_2	688	0.000123	84
$u(4_1 \cup \phi)$	7^+_6	16	0.000003	366
$u(4_1 \cup \phi)$	7^+_7	11	0.000002	440
$u(4_1 \cup \phi)$	8^+_{12}	4	0.000001	556
$u(4_1 \cup \phi)$	$3^+_1\#4_1$	732	0.000131	98
$u(4_1 \cup \phi)$	$4_1\#4_1$	19	0.000003	508
l_unk.	f	126	0.000022	1,018
l_unk.	0_1	96	0.000017	5,774
l_unk.	4_1	1,116	0.000199	188
l_unk.	$3^+_1\#4_1$	83	0.000015	5,202
l_unk.	l_unk.	1	0.000000	1,354

C.6 Knot Type 5_1^+ Tables

Table C.13: Comparing SP to IKE knot type 5_1^+

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	494,842,720	88.364771	–
f	0_1	1,167,093	0.208409	44
f	5_1^+	4,286,199	0.765393	46
f	5_1^-	192	0.000034	90
f	6_1^-	405,500	0.072411	46
f	7_2^-	365	0.000065	112
f	7_6^+	403	0.000072	80
f	8_4^+	226	0.000040	112
f	8_5^+	158	0.000028	144
f	8_{14}^-	1	0.000000	624
f	8_{20}^+	28	0.000005	148
f	8_{20}^-	253	0.000045	98
f	9_1^+	2	0.000000	314
f	9_4^-	1	0.000000	2,408
f	9_8^+	1	0.000000	486
f	9_9^+	2	0.000000	478
f	9_{16}^-	3	0.000001	276
f	10_{124}^-	1	0.000000	558
f	10_{128}^+	2	0.000000	712
f	10_{131}^-	3	0.000001	490
f	10_{138}^-	1	0.000000	582
f	10_{150}^-	2	0.000000	526
f	l_unk.	8	0.000001	290
f	6_2^+	171,682	0.030658	50
f	7_2^+	5	0.000001	638
f	8_{21}^-	416	0.000074	84
f	$3_1^+\#5_1^+$	2,759	0.000493	112
f	$3_1^-\#5_1^+$	108	0.000019	150
f	$3_1^-\#5_2^-$	34	0.000006	184
f	$4_1\#4_1$	6	0.000001	178
5_1^+	0_1	4,782	0.000854	56
5_1^+	5_1^+	48,262,754	8.618349	42
5_1^+	6_1^-	1,610	0.000287	62
5_1^+	6_2^+	775	0.000138	70
5_1^+	7_6^+	1	0.000000	1,008
5_1^+	8_5^+	2	0.000000	260
5_1^+	8_{20}^-	1	0.000000	184
5_1^+	8_{21}^-	1	0.000000	364
5_1^+	$3_1^+\#5_1^+$	10	0.000002	294
5_1^+	$3_1^-\#5_1^+$	1	0.000000	3,538
5_1^+	$3_1^-\#5_2^-$	1	0.000000	196
7_1^+	0_1	4,656,922	0.831593	44
7_1^+	5_1^+	3,393	0.000606	62

Continued on next page

Table C.13 – *Comparing SP to IKE knot type 5_1^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
7_1^+	6_1^-	204	0.000036	78
7_1^+	6_2^+	67	0.000012	76
7_3^+	0_1	218	0.000039	96
7_3^+	5_1^+	1,697	0.000303	72
7_3^+	6_1^-	2,171,053	0.387688	50
7_3^+	6_2^+	33	0.000006	90
7_5^+	0_1	117	0.000021	88
7_5^+	5_1^+	873	0.000156	72
7_5^+	6_1^-	33	0.000006	114
7_5^+	6_2^+	1,037,898	0.185339	52
8_{15}^+	5_1^+	2	0.000000	462
8_{15}^+	7_6^+	3,259	0.000582	88
8_{19}^+	0_1	2,359,146	0.421276	42
8_{19}^+	5_1^+	1,909	0.000341	62
8_{19}^+	6_1^-	73	0.000013	66
8_{19}^+	6_2^+	31	0.000006	100
8_{19}^+	8_{20}^-	1	0.000000	236
8_{19}^+	$3_1^+\#5_1^+$	2	0.000000	344
9_6^+	7_2^-	1,808	0.000323	108
9_{11}^-	6_1^-	1	0.000000	1,186
9_{11}^-	8_4^+	1,142	0.000204	126
9_{13}^-	5_1^+	1	0.000000	470
9_{13}^-	$3_1^-\#5_2^-$	266	0.000048	164
9_{23}^+	$4_1\#4_1$	75	0.000013	160
9_{36}^-	0_1	1	0.000000	544
9_{36}^-	8_5^+	813	0.000145	104
9_{38}^+	8_{14}^-	1	0.000000	284
9_{49}^-	8_{21}^-	2,661	0.000475	90
9_{49}^-	5_1^+	3	0.000001	140
9_{49}^-	6_2^+	1	0.000000	478
10_2^+	9_1^+	10	0.000002	264
10_{14}^+	9_9^+	4	0.000001	310
10_{72}^-	9_{16}^-	14	0.000002	496
10_{134}^-	5_1^+	3	0.000001	640
10_{134}^-	8_{20}^-	1,228	0.000219	88
10_{154}^-	8_{20}^+	134	0.000024	94
10_{161}^+	5_1^-	1,227	0.000219	92
$3_1^+\#5_1^+$	0_1	65	0.000012	140
$3_1^+\#5_1^+$	5_1^+	591,002	0.105536	62
$3_1^+\#5_1^+$	6_1^-	19	0.000003	80
$3_1^+\#5_1^+$	6_2^+	5	0.000001	136
unk.	0_1	1	0.000000	432
unk.	5_1^+	6	0.000001	264
unk.	7_2^+	7	0.000001	588
unk.	9_4^-	14	0.000002	1,926

Continued on next page

Table C.13 – *Comparing SP to IKE knot type 5_1^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
unk.	10_8^+	4	0.000001	476
unk.	10_{46}^-	1	0.000000	1,068
unk.	10_{124}^-	16	0.000003	198
unk.	10_{126}^-	1	0.000000	820
unk.	10_{128}^+	5	0.000001	564
unk.	10_{128}^-	1	0.000000	654
unk.	10_{131}^-	41	0.000007	248
unk.	10_{138}^-	4	0.000001	408
unk.	10_{150}^-	4	0.000001	206
unk.	10_{152}^-	2	0.000000	372
unk.	$3_1^+ \# 5_1^+$	13,826	0.002469	102
unk.	$3_1^- \# 5_1^+$	568	0.000101	138
unk.	l_unk.	35	0.000006	428

Table C.14: Comparing SP to possible DLE for knot type 5_1^+

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
5_1^+	$4_1^2 b$	1,661	0.003441	62
5_1^+	$6_1^2 a$	2,994	0.006203	56
5_1^+	$6_2^2 a$	1	0.000002	364
5_1^+	$7_4^2 c$	2	0.000004	260
5_1^+	$7_7^2 d$	1,869	0.003872	56
5_1^+	$8_2^2 b$	1	0.000002	196
5_1^+	$9_{51}^2 c$	2	0.000004	184
5_1^+	$l(3_1^+ \cup \phi)$	784	0.001624	70
5_1^+	$l(5_1^+ \cup \phi)$	550	0.001139	64
5_1^+	$l(5_1^+ \cup \phi)$	10	0.000021	294
5_1^+	$l(5_2^+ \cup \phi)$	1	0.000002	1,008
5_1^+	$u(5_1^+ \cup \phi)$	48,260,402	99.980244	42
5_1^+	l_unk.	1,661	0.003441	154

Table C.15: Comparing DLE to possible SP knot type 5_1^+

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	494,842,720	88.364771	–
$4_1^2 b$	f	405,509	0.072412	46
$4_1^2 b$	5_1^+	1,661	0.000297	62
$4_1^2 b$	7_1^+	202	0.000036	78
$4_1^2 b$	7_3^+	2,171,084	0.387694	50

Continued on next page

Table C.15 – *Comparing DLE to possible SP knot type* 5_1^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
4_1^2b	7_5^+	30	0.000005	114
4_1^2b	8_{19}^+	75	0.000013	66
4_1^2b	$9_{11}s$	1	0.000000	1,186
4_1^2b	$3_1^+\#5_1^+$	19	0.000003	80
6_1^2a	f	788,157	0.140742	48
6_1^2a	5_1^+	2,994	0.000535	56
6_1^2a	7_1^+	4,656,811	0.831573	44
6_1^2a	7_3^+	149	0.000027	96
6_1^2a	7_5^+	60	0.000011	90
6_1^2a	8_{19}^+	110	0.000020	72
6_1^2a	$3_1^+\#5_1^+$	44	0.000008	140
6_1^2a	l_unk.	1	0.000000	432
6_2^2a	f	416	0.000074	84
6_2^2a	5_1^+	1	0.000000	364
6_2^2a	9_{49}^-	2,662	0.000475	90
7_1^2a	f	226	0.000040	112
7_4^2c	f	158	0.000028	144
7_4^2c	5_1^+	2	0.000000	260
7_4^2c	9_{36}^-	813	0.000145	104
7_7^2d	f	379,207	0.067716	44
7_7^2d	5_1^+	1,869	0.000334	56
7_7^2d	7_1^+	179	0.000032	82
7_7^2d	7_3^+	63	0.000011	106
7_7^2d	7_5^+	58	0.000010	88
7_7^2d	8_{19}^+	2,359,072	0.421263	42
7_7^2d	$3_1^+\#5_1^+$	20	0.000004	184
7_7^2d	9_{36}^-	1	0.000000	544
8_2^2b	f	34	0.000006	184
8_2^2b	5_1^+	1	0.000000	196
8_2^2b	9_{13}^-	266	0.000048	164
8_2^2d	f	365	0.000065	112
8_2^2d	9_6^+	1,808	0.000323	108
8_5^2a	f	1	0.000000	624
8_5^2a	9_{38}^+	1	0.000000	284
8_5^2c	f	3	0.000001	490
8_5^2c	l_unk.	41	0.000007	248
8_6^2a	f	6	0.000001	178
8_6^2a	9_{23}^+	75	0.000013	160
8_7^2b	f	1	0.000000	582
8_7^2b	l_unk.	4	0.000001	408
9_4^2a	l_unk.	4	0.000001	476
9_{19}^2c	l_unk.	1	0.000000	1,068
9_{20}^2a	f	1	0.000000	290
9_{44}^2a	f	4	0.000001	696
9_{44}^2a	l_unk.	13	0.000002	710
9_{44}^2a	f	4	0.000001	696
9_{51}^2c	f	253	0.000045	98

Continued on next page

Table C.15 – *Comparing DLE to possible SP knot type* 5_1^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$9^2_{51}c$	8^+_{19}	1	0.000000	236
$9^2_{51}c$	5^+_1	2	0.000000	184
$9^2_{51}c$	10^-_{134}	1,228	0.000219	88
$9^2_{56}c$	f	2	0.000000	526
$9^2_{56}c$	l_unk.	4	0.000001	206
$9^2_{60}d$	f	28	0.000005	148
$9^2_{60}d$	10^-_{154}	134	0.000024	94
$9^2_{61}a$	f	192	0.000034	90
$9^2_{61}a$	10^+_{161}	1,227	0.000219	92
$l(3^+_1 \cup \phi)$	f	171,685	0.030658	50
$l(3^+_1 \cup \phi)$	5^+_1	784	0.000140	70
$l(3^+_1 \cup \phi)$	7^+_1	66	0.000012	76
$l(3^+_1 \cup \phi)$	7^+_3	31	0.000006	90
$l(3^+_1 \cup \phi)$	7^+_5	1,037,917	0.185342	52
$l(3^+_1 \cup \phi)$	8^+_{19}	30	0.000005	100
$l(3^+_1 \cup \phi)$	$3^+_1\#5^+_1$	4	0.000001	344
$l(3^+_1 \cup \phi)$	9^-_{49}	1	0.000000	478
$l(5^+_1 \cup \phi)$	f	140,606	0.025108	62
$l(5^+_1 \cup \phi)$	5^+_1	550	0.000098	64
$l(5^+_1 \cup \phi)$	7^+_1	90	0.000016	90
$l(5^+_1 \cup \phi)$	7^+_3	27	0.000005	90
$l(5^+_1 \cup \phi)$	7^+_5	9	0.000002	160
$l(5^+_1 \cup \phi)$	8^+_{19}	25	0.000004	138
$l(5^+_1 \cup \phi)$	$3^+_1\#5^+_1$	590,472	0.105441	62
$l(5^+_1 \cup \phi)$	f	2,758	0.000492	112
$l(5^+_1 \cup \phi)$	5^+_1	10	0.000002	294
$l(5^+_1 \cup \phi)$	8^+_{19}	2	0.000000	344
$l(5^+_1 \cup \phi)$	l_unk.	13,823	0.002468	102
$l(5^+_2 \cup \phi)$	f	403	0.000072	80
$l(5^+_2 \cup \phi)$	5^+_1	1	0.000000	1,008
$l(5^+_2 \cup \phi)$	8^+_{15}	32,589	0.005819	88
$l(7^+_1 \cup \phi)$	f	2	0.000000	314
$l(7^+_1 \cup \phi)$	10^+_2	10	0.000002	264
$l(7^-_3 \cup \phi)$	f	2	0.000000	478
$l(7^-_3 \cup \phi)$	10^+_{14}	4	0.000001	310
$l(7^+_5 \cup \phi)$	f	3	0.000001	276
$l(7^+_5 \cup \phi)$	10^-_{72}	14	0.000002	496
$u(5^+_1 \cup \phi)$	f	4,145,114	0.740199	46
$u(5^+_1 \cup \phi)$	5^+_1	48,260,402	8.617929	42
$u(5^+_1 \cup \phi)$	7^+_1	3,261	0.000582	62
$u(5^+_1 \cup \phi)$	7^+_3	1,647	0.000294	72
$u(5^+_1 \cup \phi)$	7^+_5	849	0.000152	72
$u(5^+_1 \cup \phi)$	8^+_{15}	2	0.000000	462
$u(5^+_1 \cup \phi)$	8^+_{19}	1,847	0.000330	62
$u(5^+_1 \cup \phi)$	$3^+_1\#5^+_1$	493	0.000088	78
$u(5^+_1 \cup \phi)$	9^-_{13}	1	0.000000	470

Continued on next page

Table C.15 – *Comparing DLE to possible SP knot type* 5_1^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$u(5_1^+ \cup \phi)$	9_{49}^-	2	0.000000	140
$u(5_1^+ \cup \phi)$	10_{134}^-	3	0.000001	640
$u(5_1^+ \cup \phi)$	l_unk.	7	0.000001	264
l_unk.	f	255	0.000046	136
l_unk.	5_1^+	1,661	0.000297	154
l_unk.	7_1^+	5	0.000001	162
l_unk.	$3_1^+ \# 5_1^+$	39	0.000007	5,450
l_unk.	l_unk.	638	0.000114	138

C.7 Knot Type 5_1^- Tables

Table C.16: Comparing SP to IKE knot type 5_1^-

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	492,608,514	87.965806	–
f	0_1	3,194,818	0.570503	44
f	5_1^+	854	0.000152	90
f	5_1^-	2,014,416	0.359717	46
f	6_1^+	3,993	0.000713	78
f	6_2^-	2,334	0.000417	80
f	7_2^-	1	0.000000	1,190
f	7_6^-	18	0.000003	174
f	8_4^-	1	0.000000	358
f	8_5^-	5	0.000001	576
f	8_{20}^+	3,804	0.000679	70
f	8_{20}^-	175	0.000031	124
f	9_3^+	1	0.000000	660
f	9_4^+	6	0.000001	478
f	9_4^-	2	0.000000	372
f	9_7^-	1	0.000000	1,130
f	10_{126}^+	6	0.000001	430
f	10_{128}^+	1	0.000000	202
f	10_{128}^-	3	0.000001	462
f	10_{133}^-	1	0.000000	400
f	10_{133}^+	1	0.000000	722
f	10_{133}^-	1	0.000000	400
f	10_{161}^+	1	0.000000	208
f	l_unk.	3	0.000001	260
f	7_2^+	6,189	0.001105	78
f	8_{21}^+	27	0.000005	220
f	$3_1^- \# 5_1^-$	55	0.000010	224

Continued on next page

Table C.16 – *Comparing SP to IKE knot type* 5_1^- *continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	$3_1^+\#5_1^-$	1,496	0.000267	112
f	$3_1^+\#5_2^+$	5	0.000001	270
f	$4_1\#4_1$	4	0.000001	414
3_1^-	0_1	31,880,370	5.692923	40
3_1^-	5_1^+	4	0.000001	138
3_1^-	5_1^-	8,307	0.001483	54
3_1^-	6_1^+	8	0.000001	174
3_1^-	6_2^-	2	0.000000	222
3_1^-	7_2^+	14	0.000002	96
3_1^-	8_{20}^+	14	0.000002	188
3_1^-	$3_1^+\#5_1^-$	4	0.000001	724
5_1^-	0_1	12,919	0.002307	60
5_1^-	5_1^+	2	0.000000	316
5_1^-	5_1^-	22,545,459	4.025975	46
5_1^-	6_1^+	13	0.000002	262
5_1^-	6_2^-	3	0.000001	206
5_1^-	7_2^+	11	0.000002	204
5_1^-	8_{20}	7	0.000001	182
5_1^-	$3_1^+\#5_1^-$	5	0.000001	322
5_2^-	0_1	7,299,159	1.303421	42
5_2^-	5_1^-	1,936	0.000346	58
5_2^-	6_1^+	3	0.000001	192
5_2^-	7_2^+	5	0.000001	146
5_2^-	8_{20}^+	2	0.000000	544
8_2^-	0_1	21	0.000004	180
8_2^-	5_1^-	13	0.000002	240
8_2^-	7_2^+	31,866	0.005690	74
8_7^-	0_1	15	0.000003	130
8_7^-	5_1^-	7	0.000001	370
8_7^-	6_1^+	20,611	0.003681	70
8_{10}^-	0_1	3	0.000001	332
8_{10}^-	5_1^-	2	0.000000	298
8_{10}^-	6_2^-	12,183	0.002176	74
8_{16}^-	8_{21}^+	129	0.000023	120
8_{21}^-	0_1	5	0.000001	172
8_{21}^-	5_1^+	7,153	0.001277	72
8_{21}^-	5_1^-	1	0.000000	356
9_{26}^+	$3_1^+\#5_2^+$	22	0.000004	122
9_{31}^-	$4_1\#4_1$	41	0.000007	174
9_{43}^+	0_1	13	0.000002	118
9_{43}^+	5_1^-	7	0.000001	338
9_{43}^+	8_{20}^+	20,087	0.003587	68
9_{45}^-	8_{20}^-	1,338	0.000239	94
10_5^+	9_3^+	3	0.000001	212
10_6^-	9_4^+	20	0.000004	314

Continued on next page

Table C.16 – *Comparing SP to IKE knot type 5_1^- continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
10_{12}^+	9_4^-	20	0.000004	442
10_{15}^+	8_4^-	26	0.000005	292
10_{54}^+	8_5^-	7	0.000001	324
10_{77}^+	9_7^-	11	0.000002	406
10_{85}^-	9_8^-	1	0.000000	690
10_{143}^-	7_2^-	9	0.000002	248
10_{151}^+	7_6^-	132	0.000024	168
$3_1^+\#5_1^-$	0_1	180	0.000032	98
$3_1^+\#5_1^-$	5_1^-	312,784	0.055854	64
unk.	l_unk.	18	0.000003	526
unk.	0_1	6	0.000001	170
unk.	5_1^-	5	0.000001	566
unk.	8_{14}	1	0.000000	254
unk.	9_1^-	1	0.000000	386
unk.	10_{126}^+	37	0.000007	228
unk.	10_{128}^+	3	0.000001	432
unk.	10_{128}^-	15	0.000003	358
unk.	10_{133}^+	2	0.000000	700
unk.	10_{133}^-	1	0.000000	294
unk.	10_{136}^+	1	0.000000	328
unk.	10_{136}^-	2	0.000000	580
unk.	10_{155}^+	2	0.000000	656
unk.	10_{160}^-	1	0.000000	612
unk.	10_{161}^+	12	0.000002	178
unk.	$3_1^+\#5_1^-$	7,820	0.001396	92

Table C.17: Comparing SP to possible DLE for knot type 5_1^-

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
5_1s	4_1^2c	8,908	0.039489	66
5_1s	6_1^2d	14	0.000062	204
5_1s	6_2^2b	2	0.000009	316
5_1s	7_1^2d	15	0.000066	262
5_1s	7_4^2a	4	0.000018	206
5_1s	7_7^2a	7	0.000031	182
5_1s	$l(3_1^- \cup \phi)$	2,217	0.009828	60
5_1s	$l(5_1^- \cup \phi)$	128	0.000567	126
5_1s	$l(5_1^- \cup \phi)$	4	0.000018	322
5_1s	$u(5_1^- \cup \phi)$	22,545,979	99.944854	46
5_1s	l_unk.	1,141	0.005058	354

Table C.18: Comparing DLE to possible SP knot type 5_1^-

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	492,608,514	87.965806	–
4_1^2c	3_1^-	31,877,553	5.692420	40
4_1^2c	5_1^-	8,908	0.001591	66
4_1^2c	5_2^-	2,542	0.000454	64
4_1^2c	8_2^-	12	0.000002	180
4_1^2c	8_7^-	14	0.000002	130
4_1^2c	8_{10}^-	2	0.000000	484
4_1^2c	8_{21}^-	4	0.000001	172
4_1^2c	9_{43}^+	10	0.000002	118
4_1^2c	$3_1^+\#5_1^-$	159	0.000028	100
4_1^2c	l_unk.	5	0.000001	170
6_1^2d	f	6,185	0.001104	78
6_1^2d	3_1^-	18	0.000003	96
6_1^2d	5_1^-	14	0.000002	204
6_1^2d	5_2^-	4	0.000001	244
6_1^2d	8_2^-	31,870	0.005691	74
6_2^2b	f	855	0.000153	90
6_2^2b	3_1^-	4	0.000001	138
6_2^2b	5_1^-	2	0.000000	316
6_2^2b	8_{21}^-	7,154	0.001277	72
7_1^2d	f	3,996	0.000714	78
7_1^2d	3_1^-	8	0.000001	174
7_1^2d	5_1^-	15	0.000003	262
7_1^2d	5_2^-	3	0.000001	192
7_1^2d	8_7^-	20,600	0.003679	70
7_4^2a	f	2,333	0.000417	80
7_4^2a	3_1^-	2	0.000000	222
7_4^2a	5_1^-	4	0.000001	206
7_4^2a	8_{10}^-	12,183	0.002176	74
7_7^2a	f	3,804	0.000679	70
7_7^2a	3_1^-	16	0.000003	188
7_7^2a	5_1^-	7	0.000001	182
7_7^2a	5_2^-	2	0.000000	544
7_7^2a	9_{43}^+	20,090	0.003587	68
8_2^2a	f	6	0.000001	478
8_2^2a	10_6^-	20	0.000004	314
8_2^2c	f	3	0.000001	462
8_2^2c	l_unk.	15	0.000003	358
8_5^2b	9_{26}^+	22	0.000004	122
8_5^2b	f	5	0.000001	270
8_5^2d	l_unk.	1	0.000000	612
8_6^2d	l_unk.	2	0.000000	580
8_7^2c	f	4	0.000001	414
8_7^2c	9_{31}^-	41	0.000007	174
9_4^2c	f	1	0.000000	358
9_4^2c	10_{15}^+	26	0.000005	292
9_{19}^2a	10_{85}^-	1	0.000000	690

Continued on next page

Table C.18 – *Comparing DLE to possible SP knot type* 5_1^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$9^2_{19}b$	f	5	0.000001	576
$9^2_{19}b$	10^+_{54}	7	0.000001	324
$9^2_{44}c$	f	1	0.000000	1190
$9^2_{44}c$	10^-_{143}	9	0.000002	248
$9^2_{51}b$	f	6	0.000001	430
$9^2_{51}b$	l_unk.	37	0.000007	228
$9^2_{56}a$	f	18	0.000003	174
$9^2_{56}a$	10^+_{151}	132	0.000024	168
$9^2_{57}c$	f	27	0.000005	220
$9^2_{57}c$	8^-_{16}	129	0.000023	120
$9^2_{61}b$	f	1	0.000000	208
$9^2_{61}b$	l_unk.	12	0.000002	178
$l(3^-_1 \cup \phi)$	f	731,378	0.130603	44
$l(3^-_1 \cup \phi)$	3^-_1	2,524	0.000451	58
$l(3^-_1 \cup \phi)$	5^-_1	2,217	0.000396	60
$l(3^-_1 \cup \phi)$	5^-_2	7,296,481	1.302943	42
$l(3^-_1 \cup \phi)$	8^-_2	4	0.000001	378
$l(3^-_1 \cup \phi)$	8^-_7	2	0.000000	174
$l(3^-_1 \cup \phi)$	8^-_{10}	1	0.000000	332
$l(3^-_1 \cup \phi)$	$3^+_1 \# 5^-_1$	37	0.000007	98
$l(3^-_1 \cup \phi)$	l_unk.	1	0.000000	256
$l5_1c$	f	71,103	0.012697	64
$l5_1c$	3^-_1	93	0.000017	100
$l5_1c$	5^-_1	128	0.000023	126
$l5_1c$	5^-_2	34	0.000006	88
$l5_1c$	$3^+_1 \# 5^-_1$	312,582	0.055818	64
$l5_1d$	f	1,497	0.000267	112
$l5_1d$	3^-_1	4	0.000001	724
$l5_1d$	5^-_1	4	0.000001	322
$l5_1d$	l_unk.	7,820	0.001396	92
$l5_2d$	f	175	0.000031	124
$l5_2d$	9^-_{45}	1,338	0.000239	94
$l7_1c$	f	1	0.000000	660
$l7_1c$	10^+_5	3	0.000001	212
$l7_3a$	f	2	0.000000	372
$l7_3a$	10^+_{12}	20	0.000004	442
$l7_5c$	f	1	0.000000	1,130
$l7_5c$	10^+_{77}	11	0.000002	406
$u5_1c$	f	1,943,231	0.347006	46
$u5_1c$	3^-_1	7,877	0.001407	54
$u5_1c$	5^-_1	22,545,979	4.026068	46
$u5_1c$	5^-_2	1,889	0.000337	58
$u5_1c$	8^-_2	14	0.000002	240
$u5_1c$	8^-_7	8	0.000001	370
$u5_1c$	8^-_{10}	2	0.000000	298
$u5_1c$	8^-_{21}	1	0.000000	356
$u5_1c$	9^+_{43}	7	0.000001	338

Continued on next page

Table C.18 – *Comparing DLE to possible SP knot type* 5_1^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$u5_1c$	$3_1^+\#5_1^-$	169	0.000030	142
$u5_1c$	l_unk.	5	0.000001	566
l_unk.	f	249	0.000044	202
l_unk.	3_1^-	624	0.000111	374
l_unk.	5_1^-	1,141	0.000204	354
l_unk.	5_2^-	150	0.000027	370
l_unk.	8_7^-	9	0.000002	6,226
l_unk.	$3_1^+\#5_1^-$	17	0.000003	5,452
l_unk.	l_unk.	410	0.000073	140

C.8 Knot Type 5_2^+ Tables

Table C.19: Comparing SP to IKE knot type 5_2^+

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	497,242,475	88.793299	36
f	0_1	946,154	0.168956	42
f	3_1^-	261,005	0.046608	50
f	4_1	229,053	0.040902	46
f	5_2^+	3,576,151	0.638598	50
f	6_1^+	413,875	0.073906	50
f	6_3^+	2,848	0.000509	84
f	7_1^-	1	0.000000	852
f	7_5^+	30	0.000005	268
f	7_5^-	169	0.000030	140
f	7_6^-	112	0.000020	168
f	7_7^-	1	0.000000	198
f	8_1^-	12	0.000002	182
f	8_6^+	251	0.000045	144
f	8_9^+	116	0.000021	184
f	8_{10}^-	2	0.000000	268
f	8_{11}^-	1	0.000000	944
f	8_{19}^+	373	0.000067	64
f	8_{21}^+	3,600	0.000643	72
f	9_7^+	4	0.000001	434
f	9_8^+	2	0.000000	352
f	9_{11}^+	1	0.000000	1,148
f	9_{42}^+	24	0.000004	188
f	10_{54}^-	1	0.000000	1,472
f	10_{127}^+	6	0.000001	372
f	10_{152}^+	5	0.000001	686

Continued on next page

Table C.19 – *Comparing SP to IKE knot type 5_2^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	l_unk.	5	0.000001	950
f	6_2^-	187,783	0.033533	48
f	8_2^+	10	0.000002	216
f	$3_1^-\#5_1^-$	18	0.000003	148
f	$3_1^+\#5_2^+$	2,399	0.000428	88
f	$3_1^-\#5_2^+$	104	0.000019	142
5_1^+	0_1	10,014,838	1.788364	42
5_1^+	3_1^-	40	0.000007	80
5_1^+	4_1	25	0.000004	88
5_1^+	5_2^+	1,767	0.000316	66
5_1^+	6_1^+	55	0.000010	72
5_1^+	6_2^-	27	0.000005	100
5_1^+	$3_1^-\#5_1^-$	1	0.000000	1,082
5_2^+	0_1	2,269	0.000405	68
5_2^+	4_1	341	0.000061	76
5_2^+	3_1^-	402	0.000072	102
5_2^+	5_2^+	40,150,347	7.169705	48
5_2^+	6_1^+	655	0.000117	88
5_2^+	6_2^-	328	0.000059	106
5_2^+	6_3^+	3	0.000001	540
5_2^+	8_{21}^+	10	0.000002	218
5_2^+	$3_1^+\#5_2^+$	5	0.000001	190
7_1^+	7_1^-	3	0.000001	418
7_2^+	0_1	60	0.000011	74
7_2^+	3_1^-	16	0.000003	186
7_2^+	4_1	14	0.000002	146
7_2^+	5_2^+	505	0.000090	70
7_2^+	6_1^+	2,540,432	0.453649	50
7_2^+	6_2^-	6	0.000001	220
7_3^+	0_1	37	0.000007	104
7_3^+	3_1^-	1,535,314	0.274163	48
7_3^+	4_1	4	0.000001	144
7_3^+	5_2^+	343	0.000061	78
7_3^+	6_1^+	13	0.000002	124
7_3^+	6_2^-	12	0.000002	318
7_4^+	0_1	37	0.000007	100
7_4^+	3_1^-	8	0.000001	212
7_4^+	4_1	6	0.000001	340
7_4^+	5_2^+	235	0.000042	80
7_4^+	6_1^+	8	0.000001	228
7_4^+	6_2^-	1,041,541	0.185989	54
7_5^+	0_1	26	0.000005	60
7_5^+	3_1^-	7	0.000001	218
7_5^+	4_1	1,274,928	0.227666	54
7_5^+	5_2^+	256	0.000046	88

Continued on next page

Table C.19 – *Comparing SP to IKE knot type 5_2^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
7_5^+	6_1^+	8	0.000001	132
7_5^+	6_2^-	3	0.000001	220
8_{15}^+	0_1	1	0.000000	628
8_{15}^+	5_2^+	1	0.000000	6,280
8_{15}^+	6_3^+	15,108	0.002698	72
9_5^-	$3_1^-\#5_1^-$	131	0.000023	132
9_{11}^-	5_1^+	78	0.000014	156
9_{12}^+	0_1	2	0.000000	2,338
9_{12}^+	5_2^+	6	0.000001	848
9_{12}^+	$3_1^+\#5_2^+$	12,291	0.002195	98
9_{15}^-	8_6^+	1,292	0.000231	120
9_{18}^+	7_5^-	770	0.000138	102
9_{20}^+	8_2^+	78	0.000014	130
9_{21}^-	8_9^+	665	0.000119	96
9_{23}^+	7_6^-	684	0.000122	122
9_{43}^-	6_1^+	1	0.000000	666
9_{43}^-	8_{19}^+	2,832	0.000506	88
9_{45}^+	0_1	1	0.000000	552
9_{45}^+	5_2^+	5	0.000001	690
9_{45}^+	6_2^-	1	0.000000	400
9_{45}	8_{21}^+	17,451	0.003116	74
10_{21}^+	9_3^-	7	0.000001	634
10_{30}^+	9_{13}^-	2	0.000000	930
10_{36}^+	9_7^+	7	0.000001	288
10_{39}^+	9_6^+	3	0.000001	316
10_{55}^+	8_{10}^-	24	0.000004	190
10_{72}^-	l_unk.	1	0.000000	246
10_{78}^+	9_8^+	3	0.000001	484
10_{92}^-	9_{20}^+	1	0.000000	274
10_{149}^+	7_5^+	182	0.000033	144
10_{157}^-	7_7^-	23	0.000004	194
10_{160}^-	9_{42}^+	99	0.000018	136
10_{165}^-	9_{45}^+	7	0.000001	1,266
$3_1^+\#5_2^+$	0_1	54	0.000010	272
$3_1^+\#5_2^+$	3_1^-	12	0.000002	360
$3_1^+\#5_2^+$	4_1	2	0.000000	474
$3_1^+\#5_2^+$	5_2^+	516,062	0.092154	64
$3_1^+\#5_2^+$	6_1^+	14	0.000002	222
$3_1^+\#5_2^+$	6_2^-	10	0.000002	436
$3_1^+\#5_1^+$	10_{125}^+	1	0.000000	864
unk.	l_unk.	23	0.000004	472
unk.	5_1^-	1	0.000000	368
unk.	8_5^-	1	0.000000	228
unk.	8_8^+	5	0.000001	536
unk.	8_{11}^-	3	0.000001	270

Continued on next page

Table C.19 – *Comparing SP to IKE knot type 5_2^+ continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
unk.	8_{14}^+	2	0.000000	312
unk.	9_{11}^+	3	0.000001	352
unk.	10_6^+	1	0.000000	438
unk.	10_{54}^-	4	0.000001	1,114
unk.	10_{124}^+	4	0.000001	164
unk.	10_{127}^+	12	0.000002	262
unk.	10_{127}^-	4	0.000001	552
unk.	10_{133}^-	5	0.000001	234
unk.	10_{148}^+	1	0.000000	396
unk.	10_{151}^+	1	0.000000	328
unk.	10_{152}^+	29	0.000005	446
unk.	$3_1^- \# 5_2^+$	463	0.000083	168

Table C.20: Comparing SP to possible DLE for knot type 5_2^+

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
5_2^+	$2_1^2 a$	246	0.000613	88
5_2^+	$4_1^2 d$	140	0.000349	106
5_2^+	$6_2^2 a$	161	0.000401	102
5_2^+	$6_3^2 b$	112	0.000279	76
5_2^+	$7_5^2 a$	5	0.000012	540
5_2^+	$7_8^2 a$	5	0.000012	726
5_2^+	$l(3_1^+ \cup \phi)$	845	0.002104	68
5_2^+	$l(5_2^+ \cup \phi)$	151	0.000376	202
5_2^+	$l(5_2^+ \cup \phi)$	4	0.000010	696
5_2^+	$u(5_2^+ \cup \phi)$	40,150,379	99.990086	48
5_2^+	l_unk.	2,312	0.005758	304

Table C.21: Comparing DLE to possible SP knot type 5_2^+

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	497,242,475	88.793299	0
$2_1^2 a$	f	413,863	0.073904	50
$2_1^2 a$	5_1^+	48	0.000009	72
$2_1^2 a$	5_2^+	246	0.000044	88
$2_1^2 a$	7_2^+	2,540,579	0.453675	50
$2_1^2 a$	7_3^+	9	0.000002	124
$2_1^2 a$	7_4^+	4	0.000001	494
$2_1^2 a$	7_5^+	6	0.000001	132

Continued on next page

Table C.21 – *Comparing DLE to possible SP knot type* 5_2^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
4_1^2d	f	187,776	0.033531	48
4_1^2d	9_{43}^-	1	0.000000	666
4_1^2d	9_{45}^+	1	0.000000	400
4_1^2d	5_1^+	24	0.000004	114
4_1^2d	5_2^+	140	0.000025	106
4_1^2d	7_2^+	6	0.000001	290
4_1^2d	7_3^+	10	0.000002	322
4_1^2d	7_4^+	1,041,607	0.186001	54
4_1^2d	7_5^+	2	0.000000	220
4_1^2d	$3_1^+\#5_2^+$	2	0.000000	436
6_2^2a	f	261,006	0.046608	50
6_2^2a	5_1^+	36	0.000006	80
6_2^2a	5_2^+	161	0.000029	102
6_2^2a	7_2^+	9	0.000002	318
6_2^2a	7_3^+	1,535,408	0.274180	48
6_2^2a	7_4^+	8	0.000001	212
6_2^2a	7_5^+	8	0.000001	218
6_2^2a	$3_1^+\#5_2^+$	5	0.000001	462
6_3^2b	f	229,049	0.040902	46
6_3^2b	5_1^+	21	0.000004	88
6_3^2b	5_2^+	112	0.000020	76
6_3^2b	7_2^+	7	0.000001	146
6_3^2b	7_3^+	3	0.000001	704
6_3^2b	7_4^+	4	0.000001	340
6_3^2b	7_5^+	1,275,002	0.227679	54
7_1^2a	f	24	0.000004	188
7_1^2a	10_{160}^-	99	0.000018	136
7_2^2b	f	116	0.000021	184
7_2^2b	9_{21}^-	665	0.000119	96
7_3^2a	f	251	0.000045	144
7_3^2a	9_{15}^-	1,292	0.000231	120
7_5^2a	f	2,848	0.000509	84
7_5^2a	5_2^+	5	0.000001	540
7_5^2a	8_{15}^+	15,107	0.002698	72
7_8^2a	f	3,600	0.000643	72
7_8^2a	5_1^+	2	0.000000	150
7_8^2a	5_2^+	5	0.000001	726
7_8^2a	9_{45}^+	17,452	0.003116	74
8_1^2d	f	18	0.000003	148
8_1^2d	9_5^-	131	0.000023	132
8_3^2c	l_unk.	4	0.000001	552
8_4^2d	f	169	0.000030	140
8_4^2d	9_{18}^+	770	0.000138	102
8_5^2a	f	112	0.000020	168
8_5^2a	9_{23}^+	684	0.000122	122
8_5^2c	l_unk.	5	0.000001	234
9_3^2a	l_unk.	1	0.000000	438

Continued on next page

Table C.21 – *Comparing DLE to possible SP knot type* 5_2^+
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$9^2_{18}c$	10^-_{92}	1	0.000000	274
$9^2_{20}a$	f	2	0.000000	268
$9^2_{20}a$	10_{55}	24	0.000004	190
$9^2_{46}a$	10^-_{165}	7	0.000001	1,266
$9^2_{50}a$	l_unk.	1	0.000000	1,160
$9^2_{52}a$	f	6	0.000001	372
$9^2_{52}a$	l_unk.	12	0.000002	262
$9^2_{55}a$	f	30	0.000005	268
$9^2_{55}a$	10^+_{149}	182	0.000033	144
$9^2_{58}c$	f	1	0.000000	198
$9^2_{58}c$	10^-_{157}	23	0.000004	194
$9^2_{59}a$	f	5	0.000001	686
$9^2_{59}a$	l_unk.	29	0.000005	446
$l(3^+_1 \cup \phi)$	f	946,041	0.168936	42
$l(3^+_1 \cup \phi)$	5^+_1	10,015,053	1.788402	42
$l(3^+_1 \cup \phi)$	5^+_2	845	0.000151	68
$l(3^+_1 \cup \phi)$	7^+_2	45	0.000008	74
$l(3^+_1 \cup \phi)$	7^+_3	29	0.000005	104
$l(3^+_1 \cup \phi)$	7^+_4	22	0.000004	100
$l(3^+_1 \cup \phi)$	7^+_5	19	0.000003	60
$l(3^+_1 \cup \phi)$	8^+_{15}	1	0.000000	628
$l(3^+_1 \cup \phi)$	$3^+_1 \# 5^+_2$	14	0.000002	272
$l(3^+_1 \cup \phi)$	9^+_{12}	1	0.000000	4,302
$l(3^+_1 \cup \phi)$	9^+_{45}	2	0.000000	552
$l(5^+_1 \cup \phi)$	f	373	0.000067	64
$l(5^+_1 \cup \phi)$	7^+_4	1	0.000000	500
$l(5^+_1 \cup \phi)$	9^-_{43}	2,832	0.000506	88
$u(5^+_2 \cup \phi)$	f	120,995	0.021606	64
$u(5^+_2 \cup \phi)$	5^+_1	22	0.000004	378
$u(5^+_2 \cup \phi)$	5^+_2	151	0.000027	202
$u(5^+_2 \cup \phi)$	7^+_2	4	0.000001	448
$u(5^+_2 \cup \phi)$	7^+_3	9	0.000002	170
$u(5^+_2 \cup \phi)$	7^+_4	4	0.000001	552
$u(5^+_2 \cup \phi)$	7^+_5	4	0.000001	274
$u(5^+_2 \cup \phi)$	$3^+_1 \# 5^+_2$	515,907	0.092126	64
$l(5^+_2 \cup \phi)$	f	2,398	0.000428	88
$l(5^+_2 \cup \phi)$	5^+_2	4	0.000001	696
$l(5^+_2 \cup \phi)$	9^+_{12}	12,296	0.002196	98
$l(7^+_2 \cup \phi)$	f	4	0.000001	434
$l(7^+_2 \cup \phi)$	10^+_{36}	7	0.000001	288
$l(7^-_3 \cup \phi)$	10^+_{21}	7	0.000001	634
$l(7^-_4 \cup \phi)$	10^+_{30}	2	0.000000	930
$l(7^+_5 \cup \phi)$	10^+_{39}	3	0.000001	316
$l(7^+_6 \cup \phi)$	f	2	0.000000	352
$l(7^+_6 \cup \phi)$	10^+_{78}	3	0.000001	484
$n(4_1 \cup \phi)$	f	12	0.000002	182
$n(4_1 \cup \phi)$	9^-_{11}	78	0.000014	156

Continued on next page

Table C.21 – *Comparing DLE to possible SP knot type 5_2^+*
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$o(3_1^+ \cup \phi)$	f	10	0.000002	216
$u(5_2^+ \cup \phi)$	f	3,455,067	0.616976	50
$u(5_2^+ \cup \phi)$	5_1^+	1,544	0.000276	66
$u(5_2^+ \cup \phi)$	5_2^+	40,150,379	7.169711	48
$u(5_2^+ \cup \phi)$	7_2^+	383	0.000068	70
$u(5_2^+ \cup \phi)$	7_3^+	250	0.000045	82
$u(5_2^+ \cup \phi)$	7_4^+	183	0.000033	80
$u(5_2^+ \cup \phi)$	7_5^+	187	0.000033	88
$u(5_2^+ \cup \phi)$	8_{15}^+	2	0.000000	568
$u(5_2^+ \cup \phi)$	$3_1^+ \# 5_2^+$	113	0.000020	220
$u(5_2^+ \cup \phi)$	9_{12}^+	2	0.000000	848
$u(5_2^+ \cup \phi)$	9_{45}^+	3	0.000001	718
l_unk.	f	338	0.000060	142
l_unk.	5_1^+	3	0.000001	532
l_unk.	5_2^+	2,312	0.000413	304
l_unk.	7_1^+	3	0.000001	418
l_unk.	7_3^+	5	0.000001	860
l_unk.	7_4^+	2	0.000000	1,230
l_unk.	$3_1^+ \# 5_1^+$	1	0.000000	864
l_unk.	$3_1^+ \# 5_2^+$	113	0.000020	1,416
l_unk.	l_unk.	510	0.000091	164

C.9 Knot Type 5_2^- Tables

Table C.22: Comparing SP to IKE knot type 5_2^-

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	489,088,566	87.337244	36
f	0_1	83,492	0.014909	50
f	3_1^+	1,277,128	0.228059	46
f	4_1	1,790,675	0.319763	50
f	5_1^+	1	0.000000	254
f	5_2^-	2,216,136	0.395739	48
f	6_1^-	4,969	0.000887	82
f	6_3^+	52,798	0.009428	56
f	7_1^+	1	0.000000	2,190
f	7_5^+	3,184	0.000569	90
f	7_5^-	228	0.000041	142
f	7_6^+	3,063	0.000547	88
f	7_7^+	19	0.000003	190
f	8_2^-	1	0.000000	702

Continued on next page

Table C.22 – *Comparing SP to IKE knot type 5_2^- continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
f	8_6^-	6	0.000001	596
f	8_9^+	1	0.000000	454
f	8_{10}^+	52	0.000009	248
f	8_{11}^+	7	0.000001	282
f	8_{14}^-	2	0.000000	648
f	8_{19}^-	33	0.000006	228
f	9_2^-	2	0.000000	406
f	9_8^-	1	0.000000	1,168
f	9_9^+	5	0.000001	342
f	9_{11}^-	5	0.000001	440
f	9_{24}^-	1	0.000000	2,754
f	10_{124}^-	2	0.000000	516
f	10_{125}^+	1	0.000000	324
f	10_{127}^+	1	0.000000	816
f	10_{137}^-	1	0.000000	804
f	l_unk.	1	0.000000	420
f	6_2^+	2,548	0.000455	84
f	8_{21}^-	77	0.000014	132
f	$3_1^+\#5_1^+$	23	0.000004	268
f	$3_1^-\#5_2^-$	60	0.000011	286
f	$3_1^+\#5_2^-$	1,572	0.000281	106
0_1	0_1	89	0.000016	80
0_1	3_1^+	16,329,898	2.916053	42
0_1	4_1	3,744	0.000669	56
0_1	5_2^-	4,575	0.000817	70
0_1	6_1^-	4	0.000001	122
0_1	6_2^+	2	0.000000	404
0_1	6_3^+	49	0.000009	90
0_1	7_5^+	3	0.000001	298
0_1	7_6^+	5	0.000001	218
0_1	$3_1^+\#5_2^-$	2	0.000000	1,182
3_1^-	0_1	123	0.000022	64
3_1^-	3_1^+	3,637	0.000649	64
3_1^-	4_1	22,934,044	4.095365	42
3_1^-	5_2^-	6,612	0.001181	64
3_1^-	6_1^-	11	0.000002	194
3_1^-	6_2^+	7	0.000001	102
3_1^-	6_3^+	85	0.000015	116
3_1^-	7_5^+	3	0.000001	132
3_1^-	7_6^+	12	0.000002	212
5_2^-	0_1	318	0.000057	154
5_2^-	3_1^+	9,047	0.001616	60
5_2^-	4_1	13,222	0.002361	64
5_2^-	5_2^-	24,888,833	4.444434	46
5_2^-	6_1^-	22	0.000004	212

Continued on next page

Table C.22 – *Comparing SP to IKE knot type 5_2^- continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
5_2^-	6_2^+	5	0.000001	764
5_2^-	6_3^+	177	0.000032	94
5_2^-	7_5^+	19	0.000003	366
5_2^-	7_5^-	3	0.000001	178
5_2^-	7_6^+	19	0.000003	274
5_2^-	$3_1^+\#5_2^-$	7	0.000001	2,190
7_6^-	0_1	7	0.000001	146
7_6^-	3_1^+	36	0.000006	210
7_6^-	4_1	89	0.000016	134
7_6^-	5_2^-	124	0.000022	138
7_6^-	6_1^-	1	0.000000	684
7_6^-	6_3^+	336,791	0.060141	58
8_8^-	3_1^+	7	0.000001	108
8_8^-	4_1	13	0.000002	358
8_8^-	5_2^-	13	0.000002	124
8_{10}^-	4_1	1	0.000000	908
8_{10}^-	7_5^-	1,808	0.000323	90
8_{11}^-	3_1^+	5	0.000001	136
8_{11}^-	4_1	4	0.000001	320
8_{11}^-	5_2^-	7	0.000001	322
8_{11}^-	7_5^+	16,483	0.002943	80
8_{13}^-	3_1^+	2	0.000000	570
8_{13}^-	4_1	2	0.000000	354
8_{13}^-	5_2^-	5	0.000001	206
8_{13}^-	6_2^+	12,901	0.002304	88
8_{14}^-	3_1^+	5	0.000001	500
8_{14}^-	4_1	2	0.000000	200
8_{14}^-	5_2^-	11	0.000002	312
8_{14}^-	7_6^+	15,364	0.002744	82
8_{16}^-	7_7^+	134	0.000024	90
8_{20}^-	0_1	519,577	0.092782	46
8_{20}^-	3_1^+	91	0.000016	96
8_{20}^-	4_1	116	0.000021	80
8_{20}^-	5_2^-	166	0.000030	106
8_{20}^-	6_3^+	3	0.000001	398
9_{12}^-	3_1^+	1	0.000000	2,342
9_{12}^-	4_1	6	0.000001	116
9_{12}^-	5_2^-	7	0.000001	718
9_{12}^-	$3_1^+\#5_2^-$	8,490	0.001516	112
9_{14}	$3_1^+\#5_1^+$	205	0.000037	178
9_{25}^-	8_{10}^+	346	0.000062	134
9_{26}	8_{11}^+	58	0.000010	138
9_{31}^-	8_{14}^-	35	0.000006	216
9_{41}^-	8_5^+	1	0.000000	320
10_{23}	9_{10}^+	6	0.000001	460

Continued on next page

Table C.22 – *Comparing SP to IKE knot type 5_2^- continued from previous page*

SP Knot Type	IKE Knot Type	Count	Percentage of Total	Smallest Such Knot
10_{24}^-	9_9^+	35	0.000006	406
10_{31}	8_9^+	6	0.000001	360
10_{34}	9_2^-	9	0.000002	304
10_{37}	8_6^-	23	0.000004	288
10_{38}^-	9_{11}^-	23	0.000004	214
10_{73}^-	9_{24}^-	3	0.000001	908
10_{87}^+	9_{36}^-	1	0.000000	856
10_{121}^-	9_{33}^-	1	0.000000	486
10_{135}^-	8_{21}^-	387	0.000069	110
10_{141}^-	7_1^+	3	0.000001	266
10_{153}^-	8_{19}^-	217	0.000039	112
$3_1^+\#5_1^+$	6_1^-	25,532	0.004559	80
$3_1^+\#5_2^-$	0_1	5	0.000001	318
$3_1^+\#5_2^-$	3_1	147	0.000026	84
$3_1^+\#5_2^-$	4_1	222	0.000040	78
$3_1^+\#5_2^-$	5_2^-	340,872	0.060870	78
$3_1^+\#5_2^-$	6_3	3	0.000001	1,024
$3_1^+\#5_2^-$	7_5	1	0.000000	758
$3_1^+\#5_2^-$	7_6	1	0.000000	636
unk.	l_unk.	10	0.000002	418
unk.	8_1^+	5	0.000001	440
unk.	8_2^-	7	0.000001	244
unk.	9_8^-	8	0.000001	1,158
unk.	9_{42}^+	2	0.000000	270
unk.	9_{45}^-	9	0.000002	360
unk.	10_{124}^-	9	0.000002	116
unk.	10_{125}^+	2	0.000000	630
unk.	10_{126}^+	1	0.000000	182
unk.	10_{137}^-	2	0.000000	960
unk.	10_{138}^-	1	0.000000	322
unk.	10_{140}^-	6	0.000001	166
unk.	10_{148}^-	5	0.000001	240
unk.	$3_1^-\#5_2^-$	280	0.000050	182

Table C.23: Comparing SP to possible DLE for knot type 5_2^-

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
5_2^-	2_1^2b	3,370	0.013528	64
5_2^-	4_1^2a	2,294	0.009209	60
5_2^-	6_2^2b	6	0.000024	366
5_2^-	6_3^2c	12	0.000048	570
5_2^-	7_2^2c	2	0.000008	764
5_2^-	7_3^2c	4	0.000016	212

Continued on next page

Table C.23 – *Comparing SP to possible DLE for knot type* 5_2^-
continued from previous page

SP Knot Type	DLE Link Type	Count	Percentage of Total	Smallest Such Knot
5_2^-	7_8^2c	80	0.000321	154
5_2^-	$l(3_1^- \cup \phi)$	62	0.000249	94
5_2^-	$l(5_1^- \cup \phi)$	1	0.000004	178
5_2^-	$l(5_2^- \cup \phi)$	116	0.000466	230
5_2^-	$l(5_2^- \cup \phi)$	2	0.000008	2,190
5_2^-	$u(5_2^- \cup \phi)$	24,903,452	99.967003	46
5_2^-	l_unk.	2,271	0.009116	268

Table C.24: Comparing DLE to possible SP knot type 5_2^-

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
f	f	489,088,566	87.337244	0
2_1^2b	f	1,790,028	0.319648	50
2_1^2b	0_1	2,374	0.000424	56
2_1^2b	3_1^-	22,937,482	4.095979	42
2_1^2b	5_2^-	3,370	0.000602	64
2_1^2b	7_6^-	49	0.000009	176
2_1^2b	8_8^-	6	0.000001	358
2_1^2b	8_{11}^-	1	0.000000	656
2_1^2b	8_{13}^-	1	0.000000	354
2_1^2b	8_{14}^-	2	0.000000	200
2_1^2b	8_{20}^-	70	0.000012	80
2_1^2b	9_{12}^-	1	0.000000	116
2_1^2b	$3_1^+\#5_2^-$	46	0.000008	78
4_1^2a	f	1,276,660	0.227975	46
4_1^2a	0_1	16,332,672	2.916549	42
4_1^2a	3_1^-	2,237	0.000399	64
4_1^2a	5_2^-	2,294	0.000410	60
4_1^2a	7_6^-	18	0.000003	220
4_1^2a	8_8^-	4	0.000001	108
4_1^2a	8_{11}^-	4	0.000001	136
4_1^2a	8_{14}^-	2	0.000000	500
4_1^2a	8_{20}^-	58	0.000010	96
4_1^2a	$3_1^+\#5_2^-$	32	0.000006	84
6_2^2b	f	3,183	0.000568	90
6_2^2b	0_1	2	0.000000	298
6_2^2b	3_1^-	2	0.000000	214
6_2^2b	5_2^-	6	0.000001	366
6_2^2b	8_{11}^-	16,488	0.002944	80
6_3^2c	f	3,066	0.000548	88
6_3^2c	0_1	7	0.000001	218
6_3^2c	3_1^-	8	0.000001	234
6_3^2c	5_2^-	12	0.000002	570
6_3^2c	8_{14}^-	15,366	0.002744	82

Continued on next page

Table C.24 – *Comparing DLE to possible SP knot type* 5_2^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
6_3^2c	$3_1^+\#5_2^-$	1	0.000000	636
7_1^2d	f	19	0.000003	190
7_1^2d	8_{16}^-	134	0.000024	90
7_2^2c	f	2,548	0.000455	84
7_2^2c	0_1	1	0.000000	866
7_2^2c	3_1^-	5	0.000001	506
7_2^2c	5_2^-	2	0.000000	764
7_2^2c	8_{13}^-	12,905	0.002304	88
7_3^2c	f	4,969	0.000887	82
7_3^2c	0_1	4	0.000001	122
7_3^2c	3_1^-	6	0.000001	194
7_3^2c	5_2^-	4	0.000001	212
7_3^2c	8_8^-	25,548	0.004562	80
7_3^2c	$3_1^+\#5_2^-$	1	0.000000	2,344
7_5^2d	f	52	0.000009	248
7_5^2d	9_{25}^-	346	0.000062	134
7_8^2c	f	83,489	0.014909	50
7_8^2c	0_1	55	0.000010	80
7_8^2c	3_1^-	82	0.000015	64
7_8^2c	5_2^-	80	0.000014	154
7_8^2c	7_6^-	6	0.000001	146
7_8^2c	8_{20}^-	519,705	0.092804	46
7_8^2c	$3_1^+\#5_2^-$	3	0.000001	318
8_1^2a	f	2	0.000000	516
8_1^2a	l_unk.	9	0.000002	116
8_3^2b	f	23	0.000004	268
8_3^2b	9_{14}	205	0.000037	178
8_4^2a	f	5	0.000001	342
8_4^2a	10_{24}^-	35	0.000006	406
8_5^2d	f	5	0.000001	440
8_5^2d	10_{38}^-	23	0.000004	214
8_9^2c	9_{41}^-	1	0.000000	320
9_3^2d	f	6	0.000001	596
9_3^2d	10_{37}	23	0.000004	288
9_5^2b	f	1	0.000000	454
9_5^2b	10_{31}	6	0.000001	360
9_{28}^2b	10_{87}	1	0.000000	856
9_{39}^2a	10_{121}^-	1	0.000000	486
9_{46}^2c	l_unk.	6	0.000001	166
9_{50}^2d	f	1	0.000000	702
9_{50}^2d	l_unk.	7	0.000001	244
9_{51}^2d	f	1	0.000000	2,190
9_{52}^2d	f	77	0.000014	132
9_{52}^2d	10_{135}^-	387	0.000069	110
9_{55}^2c	l_unk.	5	0.000001	240
9_{59}^2d	f	33	0.000006	228
9_{59}^2d	10_{153}^-	217	0.000039	112

Continued on next page

Table C.24 – *Comparing DLE to possible SP knot type* 5_2^-
continued from previous page

DLE Link Type	SP Knot Type	Count	Percentage of Total	Smallest Such Knot
$l(3_1^- \cup \phi)$	f	52,787	0.009426	56
$l(3_1^- \cup \phi)$	0_1	29	0.000005	90
$l(3_1^- \cup \phi)$	3_1^-	55	0.000010	116
$l(3_1^- \cup \phi)$	5_2^-	62	0.000011	94
$l(3_1^- \cup \phi)$	7_6^-	336,881	0.060157	58
$l(3_1^- \cup \phi)$	8_{20}^-	3	0.000001	398
$l(3_1^- \cup \phi)$	$3_1^+ \# 5_2^-$	1	0.000000	1,024
$l(5_1^- \cup \phi)$	f	228	0.000041	142
$l(5_1^- \cup \phi)$	5_2^-	1	0.000000	178
$l(5_1^- \cup \phi)$	8_{10}^-	1,809	0.000323	90
$l(5_2^- \cup \phi)$	f	78,571	0.014031	62
$l(5_2^- \cup \phi)$	0_1	45	0.000008	126
$l(5_2^- \cup \phi)$	3_1^-	67	0.000012	112
$l(5_2^- \cup \phi)$	5_2^-	116	0.000021	230
$l(5_2^- \cup \phi)$	7_6^-	2	0.000000	2,306
$l(5_2^- \cup \phi)$	8_{20}^-	2	0.000000	360
$l(5_2^- \cup \phi)$	$3_1^+ \# 5_2^-$	340,992	0.060891	78
$l(5_2^- \cup \phi)$	f	1,564	0.000279	106
$l(5_2^- \cup \phi)$	0_1	2	0.000000	1,182
$l(5_2^- \cup \phi)$	5_2^-	2	0.000000	2,190
$l(5_2^- \cup \phi)$	9_{12}^-	8,460	0.001511	112
$l(7_2^- \cup \phi)$	f	2	0.000000	406
$l(7_2^- \cup \phi)$	10_{34}	9	0.000002	304
$l(7_3^+ \cup \phi)$	10_{23}	6	0.000001	460
$l(7_6^- \cup \phi)$	f	1	0.000000	2,754
$l(7_6^- \cup \phi)$	10_{73}^-	3	0.000001	908
$n(4_1 \cup \phi)$	f	7	0.000001	282
$n(4_1 \cup \phi)$	9_{26}	58	0.000010	138
$o(3_1^- \cup \phi)$	f	2	0.000000	648
$o(3_1^- \cup \phi)$	9_{31}^-	35	0.000006	216
$u(5_2^- \cup \phi)$	f	2,138,482	0.381872	48
$u(5_2^- \cup \phi)$	0_1	3,103	0.000554	72
$u(5_2^- \cup \phi)$	3_1^-	4,464	0.000797	64
$u(5_2^- \cup \phi)$	5_2^-	24,903,452	4.447045	46
$u(5_2^- \cup \phi)$	7_6^-	89	0.000016	138
$u(5_2^- \cup \phi)$	8_8^-	6	0.000001	124
$u(5_2^- \cup \phi)$	8_{11}^-	6	0.000001	322
$u(5_2^- \cup \phi)$	8_{13}^-	4	0.000001	206
$u(5_2^- \cup \phi)$	8_{14}^-	8	0.000001	312
$u(5_2^- \cup \phi)$	8_{20}^-	114	0.000020	106
$u(5_2^- \cup \phi)$	9_{12}^-	1	0.000000	1,930
$u(5_2^- \cup \phi)$	$3_1^+ \# 5_2^-$	96	0.000017	140
l_unk.	f	284	0.000051	254
l_unk.	0_1	77	0.000014	1,122
l_unk.	3_1^-	126	0.000022	500
l_unk.	5_2^-	2,271	0.000406	268

Continued on next page

Table C.24 – *Comparing DLE to possible SP knot type* 5_2^-
continued from previous page

DLE Link Type	**SP Knot Type**	**Count**	**Percentage of Total**	**Smallest Such Knot**
l_unk.	7_6^-	3	0.000001	$2,048$
l_unk.	8_{14}^-	4	0.000001	$6,132$
l_unk.	8_{20}^-	1	0.000000	$8,168$
l_unk.	9_{12}^-	42	0.000007	$6,160$
l_unk.	$3_1^+ \# 5_2^-$	80	0.000014	$4,618$
l_unk.	l_unk.	320	0.000057	182